



UNIVERSIDADE D
COIMBRA

Wendell da Cunha Lima

GESTÃO DE DADOS DE INVESTIGAÇÃO:
ARTICULAÇÕES E PRÁTICAS PARA A PARTILHA DOS
DADOS NA UNIVERSIDADE

Tese de Doutoramento na área de Ciência da Informação orientada pela Professora Doutora Maria Manuel de Figueiredo Costa Marques Borges e coorientada pelo Professor Doutor Licínio Gomes Roque, apresentada ao Departamento de Filosofia, Comunicação e Informação da Faculdade de Letras da Universidade de Coimbra.

Setembro de 2022

Faculdade de Letras da Universidade de Coimbra

GESTÃO DE DADOS DE INVESTIGAÇÃO: Articulações e práticas para a partilha dos dados na universidade

Wendell da Cunha Lima

Tese de Doutoramento na área de Ciência da Informação orientada pela Professora Doutora Maria Manuel de Figueiredo Costa Marques Borges e coorientada pelo Professor Doutor Licínio Gomes Roque, apresentada ao Departamento de Filosofia, Comunicação e Informação da Faculdade de Letras da Universidade de Coimbra.

Setembro de 2022



UNIVERSIDADE DE
COIMBRA

Agradecimentos

... aos meus professores orientadores. À Doutora Maria Manuel Borges, cuja compreensão e estímulo puderam me fazer prosseguir com segurança acadêmica e emocional, muito me alegro por sua convivência e amizade. Ao Doutor Licínio Roque, cujas palavras certas fizeram com que as dificuldades pudessem resultar em progresso, fornecendo o apoio necessário para que eu até aqui chegasse. Meu muito obrigado a ambos.

... à minha esposa Tina, que com seu amor e companheirismo tornou essa jornada menos árdua, agradeço por estar ao meu lado em todos os momentos do percurso, por tudo que me lembro e ainda pelo que não percebo, o meu amor e a minha gratidão.

... à minha filha Sophia, cuja sabedoria e entendimento, apesar da idade, permitiu que este trabalho se tornasse um estímulo ao ver o seu sorriso, um encorajamento ao perceber seu amor e uma força por tê-la como filha, o meu obrigado e o meu amor.

... às minhas colegas de doutoramento, Ana Maria, Edina e Solange, que puderam dividir comigo suas companhias, seus conhecimentos e suas dúvidas, pela amizade, solicitude e ajuda nesses anos de estudo e trabalho, por todo apoio, obrigado.

... aos investigadores e professores do curso de doutoramento em Ciência da Informação da Universidade de Coimbra, cujo conhecimento, disponibilidade e experiência deram significativa e inestimável contribuição ao meu aprimoramento acadêmico.

... ao Deus da minha fé, que me permitiu o equilíbrio e reequilíbrio interior necessário para desenvolver esta tese, tendo como esperança alcançar uma sociedade mais justa e igualitária, assim como Ele é.

Resumo

São evidentes a importância e o valor que os dados de investigação produzidos nas universidades possuem como fonte de recursos para a ciência e para a sociedade, potencializados quando adequadamente geridos e partilhados. A gestão de dados de investigação – RDM (*Research Data Management*) é uma preocupação para as universidades, particularmente por se tratar de um ambiente complexo e heterogêneo onde os dados podem ser invisibilizados e fragmentados por investigações díspares, com uma gama multidisciplinar de elementos humanos e não humanos, que tendem a realizar práticas isoladas. Portanto necessita-se de cooperação, articulação e coordenação. Esta tese teve como objetivo propor um modelo para a RDM que possibilita uma partilha adequada dos dados, tendo por base a articulação dos atores e o uso de práticas essenciais para a gestão, no ambiente da universidade. Sob um viés exploratório e uma abordagem qualitativa, a coleta dos dados foi realizada por intermédio de entrevistas no universo das Unidades de Investigação e Desenvolvimento (UI&D) da Universidade de Coimbra (UC), tendo como base amostral os investigadores. Além da coleta empírica, realizou-se uma análise da literatura da Ciência Aberta, da Teoria da Atividade e de nove modelos de ciclos de vida, para que o ambiente da RDM na universidade fosse contextualizado; uma partilha adequada dos dados fosse caracterizada; as práticas essenciais para gestão identificadas; uma forma de articulação dos atores estabelecida; e, o nexó teórico proposto. Como principal resultado da investigação, apresenta-se o modelo para a RDM nas UI&D da UC. Neste modelo os dados são geridos na perspectiva da articulação dos atores e das práticas essenciais de gestão, ambas direcionadas à uma partilha adequada dos dados, levando em consideração as forças que influenciam este objetivo. O modelo é considerado flexível por conter uma metodologia para que a RDM seja adaptada ao ambiente, possibilitando representar as especificidades e variar em função dos contextos das UI&D, mas também relacionar o global e o local, de modo que a universidade tenha o conhecimento dos dados geridos e as UI&D obtenham o suporte institucional para realizar a RDM. Em uma perspectiva ampla, o modelo possibilita uma RDM para a universidade, articulada, flexível e realizada com práticas essenciais, resultando em uma partilha adequada dos dados de investigação.

Palavras-chave

Gestão de dados de investigação, Ciência Aberta, Teoria da Atividade, Modelos de ciclos de vida dos dados de investigação, Partilha dos dados de investigação.

Abstract

The importance and value that research data produced in universities have as a source of resources for science and society are evident, enhanced when properly managed and shared. Research Data Management (RDM) is a concern for universities, particularly because it is a complex and heterogeneous environment where data can be invisible and fragmented by disparate investigations, with a multidisciplinary array of human and non-human elements, that tend to perform isolated practices. Therefore, cooperation, articulation and coordination are needed. This thesis aimed to propose a model for RDM which enables an adequate sharing of data, based on the articulation of actors and the use of essential practices for management, in the university environment. Under an exploratory bias and a qualitative approach, data collection was carried out through interviews in the universe of Research and Development Units (R&D) of the University of Coimbra (UC), having as sample base the researchers. Besides the empirical collection, an analysis of the Open Science literature, of the Activity Theory, and of nine models of life cycles was carried out, so that the RDM environment at the university could be contextualized; an adequate data sharing was characterized; the essential practices for management were identified; a form of articulation of the actors was established and the theoretical nexus proposed. As the main result of the research, the model for the RDM in the R&D of the UC is presented. In the model, data is managed from the perspective of stakeholder articulation and essential management practices, both of which are directed towards the appropriate sharing of data, taking into consideration the forces that influence this goal. The model is considered flexible because it contains a methodology for RDM to be adapted to the environment, making it possible to represent the specificities and to vary according to the contexts of the R&D, but also to relate the global and the local, so that the university has the knowledge of the data managed and the R&D obtain the institutional support to perform RDM. From a broad perspective, the model enables an RDM for the university that is articulated, flexible, and carried out with essential practices, resulting in an adequate sharing of research data.

Keywords

Research Data Management, Open Science, Activity Theory, Research Data Life Cycle Models, Research Data Sharing.

Lista de Tabelas

Tabela 1: Objetivos da abordagem teórica	20
Tabela 2: Categorização para a coleta e análise dos dados	22
Tabela 3: Detalhamento da categorização para a coleta e análise dos dados	23
Tabela 4: Relação entre os objetivos específicos, as fontes, os instrumentos e a análise dos dados	25
Tabela 5: Relação das 38 UI&D da Universidade de Coimbra e os domínios científicos	26
Tabela 6: Publicações analisadas sobre a relação entre a RDM e a universidade	37
Tabela 7: Referências que denotam a articulação entre atores da RDM e a universidade	37
Tabela 8: Proposta de taxonomia dos dados de investigação	49
Tabela 9: Termos e palavras oriundas dos princípios e práticas da ciência aberta	74
Tabela 10: Princípios e práticas da ciência aberta agrupados por aspectos temáticos	75
Tabela 11: Categorias de comportamentos geradoras de tensões e conflitos sob o acesso aberto..	87
Tabela 12: Vantagens dos dados abertos e da reutilização para a ciência e para a sociedade	99
Tabela 13: Princípios FAIR e guias orientadoras para os dados de investigação	101
Tabela 14: O que considerar para a realização de uma partilha adequada dos dados?.....	106
Tabela 15: Modelos de ciclos de vida dos fluxos da RDM, da curadoria e dos dados de investigação .	117
Tabela 16: Descrição do conteúdo das ações que constituem o modelo Jisc	138
Tabela 17: Etapas e práticas do modelo do UKDA	142
Tabela 18: Etapas sequenciais recorrentes nos modelos de ciclo de vida	153
Tabela 19: Elementos contextuais nos modelos de ciclo de vida	157
Tabela 20: Contribuições para o desenvolvimento do modelo para a RDM	161
Tabela 21: Tipos de colaboração dos atores	232
Tabela 22: Atores que poderiam ser úteis para a colaboração com a RDM	272
Tabela 23: As definições centrais do modelo para a RDM na universidade	272
Tabela 24: Conclusões e recomendações oriundas dos dados empíricos	273

Nota: Todas as tabelas são de elaboração própria, exceto quando a fonte é mencionada.

Lista de Figuras

Figura 1: Procedimentos e técnicas empregadas	18
Figura 2: Elementos teóricos para a construção das categorias analíticas.....	22
Figura 3: O acesso aberto e os dados abertos como componentes da ciência aberta	54
Figura 4: “Chuva” de termos e palavras a partir dos princípios e práticas da ciência aberta	73
Figura 5: Exemplo de uma organização relacional dos princípios e práticas da ciência aberta	81
Figura 6: Fatores que podem repercutir na decisão do investigador na partilha em acesso aberto....	93
Figura 7: Visão geral dos elementos para a abertura e partilha dos dados em acesso aberto	102
Figura 8: Contexto e elementos para uma partilha adequada dos dados na universidade	104
Figura 9: Visão geral das convergências, diferenciações e limites entre as práticas de curadoria e para a RDM no processo de investigação, considerando o ciclo de vida dos dados	113
Figura 10: Modelo do ciclo de vida DCC – Digital Curation Centre.....	118
Figura 11: Os dados como objeto digital central da curadoria	119
Figura 12: Primeiro grupo de ações, com 4 práticas para todo o ciclo de vida	120
Figura 13: Segundo grupo de ações, com 8 práticas sequenciais para a curadoria	121
Figura 14: Terceiro grupo de ações, com 3 práticas ocasionais de curadoria.....	123
Figura 15: Modelo de ciclo de vida DDI – Data Documentation Initiative	125
Figura 16: Modelo do ciclo de vida e-Science and the lifecycle of research	127
Figura 17: Modelo do ciclo de vida da UVa - University of Virginia	130
Figura 18: Modelo do ciclo de vida DataONE	133
Figura 19: Modelo do ciclo de vida do Jisc.....	136
Figura 20: Modelo do ciclo de vida do UK Data Archive.....	141
Figura 21: Modelo do ciclo de vida do CENS.....	145
Figura 22: Integração do modelo CENS com o processo de investigação	147
Figura 23: Modelo do ciclo de vida do IWGDD	148
Figura 24: Práticas sequenciais a serem utilizadas para a RDM em uma partilha dos dados.....	156
Figura 25: Práticas essenciais a serem observadas para a RDM.....	160
Figura 26: Estrutura da RDM, tendo por eixo as práticas essenciais	164
Figura 27: Triângulo da mediação humana	168
Figura 28: Modelo de um sistema de atividade	173
Figura 29: Fases de um ciclo de transição expansiva.....	177
Figura 30: Sequência de ações do espiral expansivo	180
Figura 31: Ciclo metodológico para a transformação e adaptação da RDM	182
Figura 32: Relação conceitual entre as práticas para a RDM, a partilha dos dados e a articulação dos atores...202	
Figura 33: Visão do domínio científico e das áreas disciplinares em relação ao propósito da investigação..204	
Figura 34: Caracterização da origem, fonte e materialidade dos dados de investigação.....210	
Figura 35: Utilidade dos dados em áreas e finalidades diferentes das quais foram planeados.....213	

Figura 36: Itens constituintes do processo ou fluxo de trabalho, em ordem aleatória	224
Figura 37: Atores internos, externos ou transversais que podem influenciar à RDM.....	231
Figura 38: Ferramentas e serviços citados nas entrevistas	242
Figura 39: Documentos citados em referência às políticas de RDM utilizadas.....	254
Figura 40: Motivos e percepções para a não partilha dos dados em acesso aberto	265
Figura 41: Motivos e percepções para a partilha dos dados em acesso aberto	268
Figura 42: Modelo para a gestão dos dados dos de investigação na universidade.....	281
Figura 43: Visão geral das sete perspectivas que compõe o modelo para a RDM	284
Figura 44: As unidades de investigação para a RDM.....	285
Figura 45: Os dados de investigação para a RDM no modelo.....	286
Figura 46: As práticas essenciais para a RDM no modelo	287
Figura 47: A articulação dos atores para a RDM modelo	293
Figura 48: As forças influenciadoras no modelo	295
Figura 49: A adaptação contextual da RDM no modelo	298
Figura 50: Uma partilha adequada dos dados de investigação.	301

Nota: Todas as figuras são de elaboração própria, exceto quando a fonte é mencionada.

Sumário

INTRODUÇÃO.....	1
UM AMBIENTE EM TRANSFORMAÇÃO.....	1
PROBLEMAS E QUESTÕES DA INVESTIGAÇÃO.....	2
PREMISSAS E OBJETIVOS.....	5
PERSPECTIVA TEÓRICA E EMPÍRICA.....	7
JUSTIFICATIVAS.....	8
ORGANIZAÇÃO DOS CAPÍTULOS.....	12
1. ABORDAGEM METODOLÓGICA - MATERIAL E MÉTODO.....	14
1.1. CARACTERIZAÇÃO E PROCEDIMENTOS.....	14
1.1.1. CLASSIFICAÇÃO METODOLÓGICA.....	14
1.1.2. ETAPAS DO PROCESSO DE INVESTIGAÇÃO.....	16
1.1.3. PROCEDIMENTOS E TÉCNICAS EMPREGADAS.....	18
1.2. CAMPO EMPÍRICO - A UNIVERSIDADE DE COIMBRA.....	26
2. A GESTÃO DOS DADOS DE INVESTIGAÇÃO NA UNIVERSIDADE.....	31
2.1. A RELAÇÃO NO AMBIENTE TEMÁTICO.....	31
2.2. OS ATORES E AS ARTICULAÇÕES NO AMBIENTE.....	36
2.3. AS CARACTERÍSTICAS GERAIS DOS DADOS DE INVESTIGAÇÃO.....	45
2.4. CONSIDERAÇÕES SOBRE A GESTÃO DOS DADOS NA UNIVERSIDADE.....	50
3. CIÊNCIA ABERTA - UMA PARTILHA ADEQUADA DOS DADOS DE INVESTIGAÇÃO.....	52
3.1. A PARTILHA EM ACESSO ABERTO.....	53
3.2. CIÊNCIA ABERTA - PRINCÍPIOS E PRÁTICAS.....	58
3.3. UMA ORGANIZAÇÃO RELACIONAL TEMÁTICA.....	72
3.4. TENSÕES E CONFLITOS SOB O ACESSO ABERTO.....	83
3.5. A ABERTURA DOS DADOS DE INVESTIGAÇÃO.....	94
3.6. CONSIDERAÇÕES PARA REALIZAR UMA PARTILHA ADEQUADA DOS DADOS.....	103
4. MODELOS DE CICLOS DE VIDA - PRÁTICAS ESSENCIAIS PARA A GESTÃO DOS DADOS.....	108
4.1. PRÁTICAS DE CURADORIA E DE GESTÃO DOS DADOS.....	109
4.2. MODELOS DE CICLOS DE VIDA.....	115
4.3. SÍNTESE DOS MODELOS DE CICLOS DE VIDA.....	151
4.4. CONSIDERAÇÕES SOBRE AS PRÁTICAS ESSENCIAIS PARA A GESTÃO DOS DADOS.....	162
5. TEORIA DA ATIVIDADE - A ARTICULAÇÃO DOS ATORES.....	165
5.1. FUNDAMENTOS ESSENCIAIS DA TEORIA DA ATIVIDADE.....	166

5.2. MODELO DO SISTEMA DE ATIVIDADE	172
5.3. A APRENDIZAGEM EXPANSIVA.....	175
5.4. UMA PROPOSTA DE MODELO DE CICLO METODOLÓGICO	181
5.5. CONSIDERAÇÕES SOBRE A ARTICULAÇÃO DOS ATORES	189
6. RESULTADOS E DISCUSSÃO	191
6.1. ARTICULAÇÕES E PRÁTICAS PARA A PARTILHA DOS DADOS – NEXO TEÓRICO	191
6.2. AMBIENTE E CARACTERÍSTICAS DOS DADOS E DA INVESTIGAÇÃO	203
6.2.1. AMBIENTE E PROPÓSITO DA INVESTIGAÇÃO.....	203
6.2.2. CARACTERÍSTICAS DOS DADOS DA INVESTIGAÇÃO	210
6.3. RECURSOS PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO	221
6.3.1. PESSOAS ENVOLVIDAS.....	221
6.3.2. PROCESSOS E FLUXOS DE TRABALHO.....	223
6.3.3. COLABORAÇÕES E INFLUÊNCIAS.....	231
6.3.4. FERRAMENTAS E SERVIÇOS	241
6.3.5. POLÍTICAS, NORMAS E ORIENTAÇÕES ADOTADAS.....	253
6.4. PARTILHA DOS DADOS DE INVESTIGAÇÃO EM ACESSO ABERTO	263
6.4.1. HÁBITOS E PERCEPÇÕES PARA A PARTILHA DOS DADOS EM ACESSO ABERTO	263
6.5. CONCLUSÕES PARA A CONCEPÇÃO DO MODELO PARA A GESTÃO DOS DADOS	271
7. PROPOSTA DE MODELO PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO	277
7.1. MODELOS CIENTÍFICOS.....	277
7.2. REPRESENTAÇÃO GRÁFICA DO MODELO PARA A GESTÃO DOS DADOS	279
7.3. MODO TEXTUAL DO MODELO.....	282
7.3.1. OS CONCEITOS CENTRAIS	282
7.3.2. A VISÃO GERAL.....	284
7.3.3. O AMBIENTE: AS UNIDADES DE INVESTIGAÇÃO DA UNIVERSIDADE	285
7.3.4. OS INSUMOS BÁSICOS: OS DADOS DE INVESTIGAÇÃO	286
7.3.5. AS PRÁTICAS ESSENCIAIS: GESTÃO E PARTILHA DOS DADOS	287
7.3.6. A ARTICULAÇÃO DOS ATORES: GESTÃO E PARTILHA DOS DADOS	293
7.3.7. AS FORÇAS TRANSVERSAIS INFLUENCIADORAS	295
7.3.8. A METODOLOGIA PARA ADAPTAÇÃO CONTEXTUAL.....	298
7.3.9. A PARTILHA ADEQUADA DOS DADOS DE INVESTIGAÇÃO	301
CONCLUSÕES.....	303
REFERÊNCIAS	330
APÊNDICES.....	355

INTRODUÇÃO

Esta tese aborda a articulação dos atores e as práticas essenciais para a gestão e a partilha adequada dos dados de investigação, realizadas no ambiente da universidade. As justificativas para tal abordagem serão sumarizadas e contextualizadas nesta introdução de forma a expor, o ambiente e os problemas vivenciados, as premissas elaboradas e os objetivos que respondem às questões levantadas, os eixos de estudo e a perspectiva empírica de abordagem, bem como a importância do trabalho, as motivações do autor e a lógica organizacional dos capítulos.

UM AMBIENTE EM TRANSFORMAÇÃO

Os dados de maneira geral têm sido alvo de debates e atenção em vários campos de atuação da ciência e da sociedade, tanto pelo grande volume existente quanto pelo valor que podem representar. É perceptível a sua valorização quando se verifica a afirmação dos editores do jornal *The Economist* (2017) ao escreverem: “The world’s most valuable resource is no longer oil, but data”. No sentido dimensional, leva-se em consideração a afirmação recorrente de Christine Borgman (2012, 2015, 2018) que um “data deluge” chega em ritmo acelerado devido aos avanços e a difusão da tecnologia digital no decorrer de várias décadas, resultando em uma produção massiva de dados digitais. Consequentemente, a influência dos dados está a transformar a ciência e a sociedade contemporânea, fazendo com que sejam utilizados para fins inovadores e em usos não previstos anteriormente.

Especificamente no campo da ciência, o volume e a importância dos dados têm aumentado significativamente. Hey, Tansley e Tolle (2009) já faziam referência à visão de Jim Gray (2007) sobre uma *e-Science*, mencionada como o quarto paradigma da ciência, afirmando não ser mais possível praticar a investigação sem que seja direcionada aos dados e ao uso de tecnologias da informação. Shearer (2015) ratifica tal entendimento ao observar que no século XXI a investigação é cada vez mais orientada aos dados. Borgman (2018) e Katz et al. (2018), acrescentam que o aumento do volume dos dados de investigação em todo o mundo, faz com que as práticas de investigação cresçam de forma célere, proporcionando novos campos de estudos, novos produtos, métodos, modelos e serviços,

influenciando uma mudança de cultura da investigação nas descobertas científicas e nos *insights* filosóficos. Portanto, essa ciência intensiva em dados pode impulsionar as descobertas, oferecer esforços concentrados e uma inovação acelerada (Wallis et al., 2013).

A gestão dos dados de investigação, denominada internacionalmente por RDM (*Research Data Management*), juntamente com a partilha dos dados em acesso aberto, vem sendo adotada de modo crescente nas instituições acadêmicas e fomentadas por agências financiadoras em todo o mundo (Shearer, 2015). A base para tal adoção está centrada no entendimento de que os dados, quando abertos e devidamente gerenciados, constituem uma fonte inestimável de recursos informacionais para a investigação científica e para a sociedade. Em consequência, a investigação acadêmica está em transformação, as práticas têm evoluído rapidamente e os dados, gerados em volumes cada vez maiores, se mostram não somente como parte do processo da investigação, mas como produtos a serem geridos e partilhados (Borgman, 2018; Katz et al., 2018).

As instituições de ensino superior, particularmente as universidades, na qualidade de entidades propulsoras da investigação e produtoras de grandes quantidades de dados, estão em meio a estas transformações e podem usufruir do envolvimento direto na temática, pois dependendo de como os dados são explorados e partilhados, podem melhor servir a sociedade, colaborar com a ciência de forma mais eficiente e, através de uma gestão adequada, viabilizar investimentos e parcerias para a aplicação na área. Ou seja, o modo como as universidades abordam a RDM e a partilha dos dados, afeta os recursos, a capacidade de empreender a investigação, os avanços sociais e científicos e o desenvolvimento da instituição em longo prazo (Sayão & Sales, 2015; Shearer, 2015).

PROBLEMAS E QUESTÕES DA INVESTIGAÇÃO

Em meio as transformações no ambiente e na própria investigação, um dos problemas verificados é que a rapidez dessas mudanças pode fazer com que os dados de investigação não sejam observados com a devida importância em todo o seu escopo e dimensões, resultando quase sempre, em serem absorvidos silenciosamente pelo tempo, pela obsolescência tecnológica e pela fragilidade das mídias digitais, de modo a perder-se em meio à sua produção massiva (Borgman, 2018; Sayão & Sales, 2014). Soma-se a isto

que as abordagens para a RDM ainda são imaturas em grande parte das universidades, pois estas não as veem, de forma clara, como pertencentes aos seus mandatos, o que resulta em uma lentidão para o envolvimento na temática (Shearer, 2015). Porquanto, ainda que reconheçam o valor dos dados e a necessidade da gestão e da partilha, a percepção é de que existe mais interesse do que investimento ou implementação nas universidades, uma vez que a maior parte das infraestruturas e práticas de investigação ainda precisam ser desenvolvidas, de modo a possibilitar uma gestão eficaz dos dados gerados pela instituição (Borgman & Bourne, 2022; Katz et al., 2018). Com isso, o que se observa é uma grande produção de dados fragmentados, “invisíveis” e práticas isoladas para a RDM, que necessitam de cooperação, articulação e coordenação (Jones, Pryor & Whyte, 2013; Whyte & Jonathan, 2011).

O problema das práticas isoladas, da fragmentação e da invisibilidade dos dados, parece ser inerentes às universidades, pois estas produzem grandes volumes de pequenas coleções de vários tipos de dados, coletados e gerados de diferentes formas, em uma infinidade de laboratórios e unidades de investigação, são dados com alto valor informacional para todas as áreas do conhecimento. Estes dados geralmente são mantidos em arquivos de computadores individuais e servidores, acompanhados por um conjunto mínimo ou inexistente de metadados e documentação. Embora sejam valiosos, os dados neste ambiente são difíceis de descobrir e utilizar, por conseguinte, recaem quase sempre no desuso e na obscuridade (e-Infrastructure Reflection Group [e-IRG], 2016; Hedstrom & Myers, 2017).

O ambiente da RDM na universidade compreende parte desta problemática, pois é constituído por um contexto complexo e heterogêneo que necessita associar profissionais da informação, investigadores, infraestruturas tecnológicas, políticas, comunidades, gestores e órgãos científicos para atuar de forma conjunta na RDM e na partilha dos dados. Soma-se a isto, que as pessoas envolvidas precisam possuir habilidades para adaptar-se as transformações e aos novos tipos de infraestruturas e serviços. Ou seja, são aspectos técnicos e sociológicos, sob uma gama qualificada e multidisciplinar de atores que, reunidos no ambiente da RDM, tendem a propagar tensões e conflitos sob a influência deste contexto multifacetado. Isto assevera a necessidade de uma estrutura bem delineada para articular atores diversos, com interesses distintos, em um esforço colaborativo que permita o uso de práticas essenciais para a gestão e a partilha adequada dos dados resultantes das investigações.

Os desafios para articular os atores, a tendência à realização de práticas isoladas na RDM, juntamente com a fragmentação e a invisibilidade dos dados, são problemas emergentes para a gestão e a partilha deste objeto informacional, acrescido de que a pressão para uma partilha pública vem de muitos quadrantes, como as agências de financiamento, editores de revistas, órgãos políticos, além das academias nacionais, conselhos de investigação e dos próprios investigadores (Borgman, 2012; Kozlov, 2022). Essas partes interessadas têm suas próprias razões para exigirem ou incentivarem tal partilha. No entanto, o ato de depositar os dados em um ambiente onde estarão disponíveis de modo aberto, não garante que possam ser explorados adequadamente, pois apesar de disponíveis nem sempre podem ser reutilizados. A possibilidade da reutilização é essencial tanto para confirmar, quanto para realizar novas descobertas no contexto científico e para o uso em benefício da sociedade, sem tal possibilidade, a partilha e todo o processo da RDM torna-se insuficiente ou inadequado.

Neste cenário complexo e heterogêneo de atores diversos, práticas isoladas para a RDM, e de dados fragmentados e invisíveis, é importante que as universidades busquem instrumentos para que a RDM, e todo o ambiente que a permeia, possa ser compreendida e organizada de modo a resultar em ações práticas e articuladas para a gestão e para a partilha dos dados, tornando-os acessíveis e disponíveis para a reutilização, potencializando assim o seu alto valor científico e social. Logo, a questão aqui de abordagem é a seguinte: de que modo uma RDM pode ser estruturada para uma partilha adequada dos dados no ambiente heterogêneo da universidade, tendo por base a articulação¹ dos atores e práticas² essenciais para a gestão dos dados? Esta questão gera as seguintes perguntas complementares: Quem são os atores da RDM? O que é uma partilha adequada dos dados? Quais são as práticas essenciais? Como deve ser a articulação dos atores? Como a RDM pode adaptar-se ao ambiente complexo e heterogêneo da universidade?

¹ O termo “articulação” é aqui utilizado em uma perspectiva semelhante ao dos sistemas, ou seja, é o que une ou liga duas ou mais partes direcionadas a um objetivo comum e onde todo estímulo, em qualquer parte, afetará às demais unidades (Bertalanffy, 1973).

² As práticas são percebidas no presente trabalho como uma forma de ação humana, realizada na tentativa de alcançar os objetivos da atividade que está a ser realizada (Ilyenkov, 1982; Leont'ev, 1978).

PREMISSAS E OBJETIVOS

A questão apresentada e os objetivos que se seguem, foram construídos com base nos estudos preliminares levando em consideração três premissas básicas. A primeira premissa está fundada na necessidade da articulação entre os atores da RDM no âmbito da universidade. Essa articulação, de modo lógico e organizado entre os atores é essencial na comunicação para gestão e para a partilha dos dados, a sua falta ou fragilidade tem como consequência uma RDM limitada e de práticas isoladas. Porém, como o ambiente em questão é complexo e heterogêneo, sem desconsiderar as especificidades, tal articulação precisa ter uma estrutura padrão, um modelo para servir de referência e vincular os atores de acordo com o contexto, possibilitando gerir e partilhar os dados adequadamente. Neste entendimento preliminar, a partilha envolve o acesso aberto e a reutilização dos dados.

A segunda premissa está centrada no fato de que, além da articulação entre os atores da RDM, é necessário o uso de práticas essenciais para gerir os dados de investigação e os partilhar de modo adequado. São práticas mínimas constituídas para o uso e sob uma estrutura padrão, observadas a complexidade técnica, humana e social do contexto. Sabe-se pela literatura que a inobservância ou a falta de práticas apropriadas para a gestão dos dados, tendem a gerar problemas de ordem tecnológicas e humanas, permitindo que influenciem de forma efetiva para a RDM e para a partilha dos dados. Entende-se, a princípio, como práticas essenciais, aquelas realizadas por atores humanos no processo da investigação, que resultam em uma partilha adequada e a consequente reutilização dos dados.

A terceira e última premissa surge da necessidade dos dados de investigação serem partilhados em acesso aberto³. A abertura dos dados e o acesso aberto constituem uma série de estratégias, fundamentadas nos princípios e práticas da ciência aberta, que compatibilizam esforços para remover as barreiras e desenvolver as ações para que seja realizada a partilha da informação científica. A partilha dos dados em acesso aberto no contexto da RDM faz com que este objeto informacional, produto da investigação, esteja público e acessível de modo que sejam reutilizáveis.

³ Os dados devem ser tão abertos quanto possível e tão fechados quanto necessário, de acordo com o estabelecido pela Comissão Europeia nesta matéria (OCDE, 2016).

As três premissas geram a hipótese de que uma RDM estará mais próxima de uma partilha adequada dos dados, se estiver organizada segundo um modelo que articule os atores e utilize práticas essenciais para gerir os dados, no contexto da universidade. Portanto, é sob o universo temático da RDM na universidade que os eixos de estudo – a partilha dos dados de investigação, a articulação dos atores e as práticas para a RDM –, serão trabalhados de modo a possibilitar que, para o ambiente da universidade, os atores sejam articulados, as práticas sejam essenciais para a RDM e a partilha dos dados seja adequada.

Considerando o universo temático e os eixos de estudo evidenciados, bem como as questões levantadas e as premissas do trabalho preliminar, o objetivo geral da investigação tem como finalidade a proposição de um modelo para a RDM, que possibilite uma partilha adequada dos dados, tendo por base a articulação dos atores e o uso de práticas essenciais para a gestão, no ambiente da universidade. Este objetivo geral tem nas Unidades de Investigação e Desenvolvimento (UI&D) da Universidade de Coimbra (UC) o seu campo empírico. De modo específico, têm-se os seguintes objetivos:

1. Contextualizar, fundamentado na literatura, o ambiente temático da RDM na universidade, identificando os atores do contexto e os dados de investigação;
2. Caracterizar uma partilha adequada dos dados de investigação, através dos fundamentos da ciência aberta;
3. Identificar as práticas que sejam essenciais para a gestão dos dados, apoiada por modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, incluindo seus elementos e processos;
4. Estabelecer uma proposta para a articulação dos atores e a adaptação contextual para a RDM, viabilizado pelos fundamentos da teoria da atividade;
5. Propor umnexo teórico entre a articulação dos atores, as práticas essenciais para a RDM e a partilha adequada dos dados de investigação; e
6. Perceber, com base na compreensão de investigadores de UI&D da UC, o ambiente e os dados de investigação para a RDM, o relacionamento dos atores e as práticas envolvidas, bem como os hábitos para a partilha dos dados.

Para alcance do produto desta tese, nos quatro primeiros objetivos, explora-se a literatura em busca da fundamentação e do contexto teórico. No quinto objetivo é fornecido a lógica relacional entre os eixos de estudo. No sexto, descreve-se e destaca-se o campo de estudo empírico na UC. E por fim, examina-se e confronta-se os produtos originados nos objetivos anteriores, permitindo o cumprimento do proposto no objetivo geral.

PERSPECTIVA TEÓRICA E EMPÍRICA

A perspectiva teórica foi estabelecida com base na revisão da literatura, efetuada a partir de pesquisa bibliográfica, com a finalidade de firmar os conceitos basilares e permitir a compreensão da relação entre a RDM e a universidade, e depois, por meio de trabalho referencial, abordar a partilha adequada dos dados de investigação e a articulação dos atores através dos fundamentos firmados na ciência aberta e na teoria da atividade. A perspectiva teórica também permitiu, mediante a análise de modelos de ciclos de vida, que representam os fluxos da RDM, da curadoria e dos dados de investigação, estabelecer as práticas essenciais para a RDM de acordo com os objetivos deste trabalho. A relação conceitual entre os eixos de estudo aqui abordados, também foi obtida nesta perspectiva. Assim, a ciência aberta, a teoria da atividade, os modelos de ciclo de vida e a relação conceitual resultante, foram os ferramentais teóricos utilizadas para conceber o modelo para a RDM proposto nesta tese.

Especificamente, a ciência aberta foi utilizada como base de todo o estudo, mas principalmente e de forma direcionada, para caracterizar uma partilha adequada dos dados de investigação. Assim, orientados pelos seus princípios e práticas, a ciência aberta permitiu definir o que seja uma partilha adequada e produziu a base para a realização de uma síntese propositiva, do que deve ser considerado e de que modo deve ser realizado uma partilha de dados em acesso aberto, pertinente ao ambiente da universidade.

A abordagem realizada através dos modelos de ciclos de vida, que representam os fluxos da RDM, da curadoria e dos dados de investigação, foi fundamental para a concepção das práticas essenciais para a gestão dos dados constantes no modelo aqui apresentado. Portanto, os modelos de ciclos de vida estudados, foram úteis para viabilizar a identificação e a proposição de um fluxo de práticas essenciais para a RDM, que deve ser observado para a realização de uma partilha adequada dos dados.

A utilização da teoria da atividade teve duas finalidades: a) fornecer a base conceitual para a concepção de uma estrutura articulada que permitiu agrupar os atores da RDM; e, b) atuar como uma lente metodológica para adaptar a RDM ao ambiente da investigação. Estas finalidades serviram tanto para contribuir com a resolução do problema da complexidade, quanto para minimizar a dificuldade causada pela heterogeneidade que afeta a RDM no âmbito da universidade. Assim, a teoria da atividade produziu uma estrutura para a articulação dos atores da RDM nas UI&D e parâmetros para a adaptação contextual.

A relação conceitual entre os eixos de estudo, tendo por base a ciência aberta, a teoria da atividade e os modelos de ciclos de vida, proporcionaram estabelecer o nexos para a abordagem teórica aqui realizada, ao mesmo tempo que resultou no modelo conceitual desta investigação, fundamentando o modelo proposto para a RDM na universidade.

A perspectiva empírica do trabalho foi realizada na UC – uma entidade com mais de sete séculos de história, que através do Instituto de Investigação Interdisciplinar (III), viabilizou a realização de entrevistas de forma estratégica, tendo as UI&D como o ambiente e o ponto central da investigação. Coube ao estudo no campo empírico, obter os dados sobre o ambiente, os atores envolvidos na RDM e os relacionamentos influentes, bem como as práticas e as percepções dos investigadores sobre a RDM e a partilha dos dados em acesso aberto. O resultado desta perspectiva permitiu o confronto com a teoria, contribuindo para a constituição do modelo para a RDM na universidade, objetivo geral do trabalho.

JUSTIFICATIVAS

A pertinência em apresentar um modelo para a RDM, que possibilita uma partilha adequada dos dados, através da articulação dos atores e de práticas essenciais para esta gestão, no ambiente da universidade, torna-se necessário pelo momento peculiar vivenciado na comunidade científica, incluindo as universidades. Por um lado, observa-se que o contexto da investigação está em transformação e o modo com que os investigadores realizam suas atividades têm mudado significativamente, direcionados pelo volume, valor e incentivos para que os dados sejam partilhados em acesso aberto. Por outro lado, percebe-se que o sistema que administra e envolve os dados nas universidades não tem acompanhado na mesma velocidade as necessidades decorrentes de tais mudanças. Portanto, é necessário

compreender este fenômeno e propor, cientificamente, alternativas que colaborem para coordenar os processos e as práticas para a RDM, com a articulação dos atores envolvidos na gestão e na partilha dos dados de investigação, observados no contexto da universidade.

No sentido acima exposto, torna-se pertinente apresentar um modelo original para a RDM, desenvolvido a partir do estudo e da análise da literatura, bem como dos dados empíricos coletados nas UI&D da UC, sob o mote conceitual da ciência aberta, da teoria da atividade e dos modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, combinando a articulação dos atores e as práticas para a gestão e a partilha dos dados, a ser realizada no ambiente da universidade.

Além do exposto, a relevância e a novidade que justificam esta investigação em nível de doutoramento, também podem ser vistas nas colaborações teóricas e práticas. Em termos teóricos, o estudo colabora como um ponto de partida para eventuais trabalhos em temáticas semelhantes, ou como uma seção de um tema novo e amplo para a área multidisciplinar da ciência da informação. Logo, não se tem a pretensão de apresentar um trabalho rígido, inflexível e definitivo, mas um avanço no conhecimento da área para a utilização em perspectivas atuais e futuras, uma vez que os limites e as interseções entre os relacionamentos dos atores, as práticas para a gestão e a partilha dos dados de investigação que este trabalho propõe, precisam ser diversamente exploradas.

Portanto, um acréscimo para o campo teórico a ser reconhecido, é que este trabalho pode contribuir com tópicos de investigação, existentes e a serem desenvolvidos na área da ciência da informação, isto se considerada a perspectiva consensual de Borko (1968), onde o autor afirma que a ciência da informação é uma disciplina que investiga as propriedades e o comportamento, as forças que governam o fluxo da informação e os meios de processamento para o seu acesso e uso de forma otimizadas, e ainda segundo o autor, a área diz respeito ao corpo de conhecimento ligado à origem, coleta, organização, armazenagem, recuperação, interpretação, transformação e utilização da informação.

Sob a perspectiva apresentada, o trabalho aqui realizado contribui para o estudo no campo da ciência da informação, visto que os dados na temática estudada possuem propriedades e comportamentos próprios, regidos por um contexto social e cultural

instituídos, que emanam da própria comunidade científica e da sociedade. Tais características exercem influência sobre os meios de processamento, acesso e uso da informação entre as partes interessadas. Nesse sentido, a definição de Borko pode ser visualizada sob diversos ângulos para este estudo da RDM na universidade.

Em suma, o acréscimo ao corpo de conhecimento da área tem a sua originalidade na construção teórica de um novo modo de compreender e tratar a RDM na universidade. Na realidade, a construção teórica será o resultado da associação de áreas e abordagens distintas, nomeadamente, a ciência aberta, a teoria da atividade e modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação. Assim, o avanço no conhecimento teórico da área pode ser percebido através das indicações de soluções para os problemas observados e confrontados com a realidade.

No aspecto da colaboração prática, tem-se que na atualidade, as universidades e as UI&D são desafiadas a adaptarem-se a uma nova realidade da investigação e administração dos dados gerados, para poderem lidar com a complexidade dos desafios informacionais emergentes. É na compreensão do contexto aqui proposto que se torna possível, de forma prática, direcionar os esforços para a consolidação da capacidade em ciência, acompanhados de investimentos que envolvem os recursos de informação para a gestão dos dados na universidade. Porém, é necessário que tais esforços e investimentos sejam realizados de modo gradual, planeado e coordenado, considerando as transformações em curso e a complexidade da atividade. Neste sentido, os resultados podem colaborar para a estruturação da RDM na instituição estudada, uma vez que propicia a verificação de uma situação atual, para que a articulação e as práticas dos participantes, com a consequente partilha dos dados de investigação, sejam realizadas de forma organizada e de modo a trazer os benefícios decorrentes para a universidade e para a sociedade.

Em relação à universidade, os benefícios podem ser viabilizados pelo aumento da eficiência e da eficácia institucional, na otimização dos recursos humanos, informacionais, infraestruturais, económicos e educacionais. Para a sociedade, a medida em que os recursos da RDM são otimizados e os dados partilhados de forma adequada em acesso aberto, o impacto social aumenta na proporção da transparência e na possibilidade da reutilização dos dados para multiformes benefícios sociais. Assim, a implementação de estratégias para a

RDM que levem em conta os atores e suas articulações, sob práticas que influenciam a partilha dos dados em acesso aberto, no ambiente da universidade, torna-se imperativo.

Quanto às motivações que justificam a escolha da temática e a elaboração desta investigação, surgem em primeiro lugar pela experiência pessoal e profissional na área de sistemas e gestão da informação. Em tal experiência, na qual o objetivo macro foi afastar pessoas socialmente vulneráveis da situação de extrema pobreza, percebeu-se que a falta de uma articulação multidisciplinar e da leitura técnico-social adequada, resultava em graves problemas na análise dos dados, interpretação da informação e tomadas de decisão, bem como a percepção de que as interferências dos atores envolvidos contribuíram para a alteração do ambiente e dos resultados pretendidos, ou seja, a falta de uma gestão organizada, que pudesse tratar tanto o processo quanto os dados na interlocução com os atores e o ambiente, gerou uma falta de confiabilidade, motivo pelo qual ocasionava restrições e receios operacionais nas ações de intervenção para a solução dos problemas.

Outra motivação determinante para a escolha da temática e elaboração da presente tese, procede da percepção de que as universidades em suas unidades de investigação, têm práticas diversas para gerir os dados produzidos pelos investigadores. Tais práticas podem ser consideradas adequadas ou inadequadas para a partilha dos dados produzidos, dependendo dos parâmetros utilizados. Nesse sentido, é importante conhecer a atividade da RDM nas UI&D de uma universidade, por isso, interessou verificar o estado e as relações envolvidas na RDM realizada em unidades de investigação da Universidade de Coimbra.

Esta tese, além de prover uma construção teórica original e avanço no corpo de conhecimento sobre a temática na área, também oferece insumos à compreensão e solução de problemas reais para a RDM a partir das UI&D na universidade. As unidades de investigação e a universidade estudada, assim como as similares a elas, também tem à disposição, através e sob a luz da construção teórica, uma proposta para o desenvolvimento de soluções direcionadas aos problemas que envolvem a articulação de atores, os dados de investigação, sua gestão e partilha de forma adequada, que podem ser utilizados em benefício de um amplo contexto da investigação na universidade, com repercussão social.

ORGANIZAÇÃO DOS CAPÍTULOS

O presente trabalho está dividido em sete capítulos. O primeiro capítulo expõe a abordagem metodológica, com a caracterização e a classificação da investigação, o detalhamento dos procedimentos adotados através do material e do método empregado, juntamente com a descrição do campo empírico na Universidade de Coimbra.

O capítulo 2 é dedicado a conhecer e contextualizar teoricamente o ambiente temático da RDM na universidade, para fundamentação inicial e desenvolvimento do estudo. Através de uma visão relacional do ambiente e com base na revisão da literatura e pesquisa bibliográfica, são identificados e classificados os atores na relação com a RDM na universidade e exibidos as características gerais dos dados de investigação. O capítulo termina com uma síntese do estudo. É delineado aqui o primeiro objetivo da tese.

O capítulo 3 refere-se ao segundo objetivo do estudo, a caracterização de uma partilha adequada dos dados de investigação. Fundamentado na ciência aberta o capítulo é dividido em seis seções. A primeira seção destaca o acesso aberto (AA) como um dos pilares da ciência aberta, são enfatizadas a importância e as características para realização da partilha em AA. Na segunda seção, reúne-se as referências sobre os princípios e práticas da ciência aberta direcionadas à partilha em AA. Na terceira seção, estrutura-se as referências anteriores e propõe-se uma organização relacional temática para partilhar os dados. Na quarta seção, sob a forma de tensões e conflitos, ressalta-se as influências cognitivas e contextuais para a partilha dos dados. A quinta seção exibe uma visão geral dos elementos para a abertura dos dados. Por último, o capítulo é sintetizado naquilo que deve ser considerado para uma partilha adequada dos dados de investigação na universidade.

O capítulo 4 é desenvolvido ao longo de quatro seções, nele são identificados e descritos as práticas essenciais para a gestão, sob a partilha dos dados de investigação, através do estudo de modelos de ciclos de vida. Para efetivar o objetivo 3 deste trabalho, na primeira seção obtém-se uma visão geral das diferenciações, convergências e limites entre as práticas de curadoria digital e da RDM para análise dos modelos. Na segunda seção são apresentados e analisados nove modelos levantados na literatura. Na terceira seção as etapas

dos modelos são sintetizadas e as práticas essenciais são identificadas e caracterizadas. Na última seção, expõe-se as considerações sobre o estudo realizado no capítulo.

O capítulo 5 se dispõe a cumprir o quarto objetivo da tese, propondo uma estrutura para articular os atores e adaptar a RDM ao ambiente. Fundamentado na teoria da atividade, o capítulo está dividido em cinco segmentos. O primeiro exhibe os fundamentos essenciais da teoria da atividade, o segundo descreve o modelo do sistema de atividade e o terceiro caracteriza a teoria da aprendizagem expansiva. Como resultado dos três primeiros, no quarto segmento do capítulo é proposto um modelo de ciclo metodológico para articulação dos atores e adaptação da RDM ao ambiente. O capítulo termina com as considerações sobre o estudo.

O capítulo 6 é dedicado à discussão dos resultados da investigação. Orientado pelos objetivos 5 e 6 deste trabalho, primeiramente apresenta-se o nexos conceitual entre os eixos de estudo – a partilha dos dados de investigação, a articulação dos atores e as práticas para a RDM. Em seguida são mostrados e analisados os dados obtidos no campo empírico da universidade através das categorias determinadas no estudo. São discutidos os resultados em cada categoria e realizadas as conclusões para a concepção do modelo para a RDM.

No capítulo 7, apresenta-se o modelo para a RDM proposto no objetivo geral do trabalho, desenvolvido a partir das ferramentas apresentadas nos capítulos anteriores, composto em essência pelas perspectivas da articulação dos atores da RDM, das práticas essenciais para a gestão e da perspectiva da partilha adequada dos dados. Os dados de investigação e o ambiente das UI&D na universidade também são perspectivas do modelo, junto com as forças transversais influenciadoras e a metodologia para a adaptação contextual. Segue-se ainda uma interpretação textual do modelo.

A tese é finalizada com as conclusões, onde os objetivos alcançados são apresentados, bem como as contribuições decorrentes do resultado do trabalho e as limitações percebidas. As conclusões são encerradas com as questões ainda em aberto e que podem ser abordadas em trabalhos futuros.

1. ABORDAGEM METODOLÓGICA – MATERIAL E MÉTODO

A construção de uma tese também deve ter como princípio servir aos outros (Eco, 1977). Portanto, para que outras pessoas compreendam este estudo, a abordagem metodológica aqui desenvolvida relata o percurso seguido metodicamente para responder à questão levantada e alcançar o objetivo deste trabalho. No desenho da metodologia que se segue, a proposta central foi realizar uma construção teórica para o confronto com a realidade. Assim, as definições metodológicas encontram-se aqui descritas em duas seções.

A primeira seção apresenta a caracterização da investigação e o detalhamento dos procedimentos empregados, através dos métodos e do material utilizado, de modo que as abordagens teórica e empírica, bem como as condutas operacionais se tornem evidentes e justificados à luz dos objetivos declarados. Já na segunda seção, é caracterizado o campo da abordagem empírica que defronte da construção teórica, resulta no produto deste estudo.

1.1. CARACTERIZAÇÃO E PROCEDIMENTOS

Conforme Marconi e Lakatos (2003), o método da investigação é o conjunto de atividades sistemáticas e racionais que permite alcançar um objetivo traçado. No entanto, tais atividades também precisam ser contextualizadas à luz de uma caracterização da investigação, para que fique mais claro como e quais foram os procedimentos realizados. Nesse sentido, segue a classificação metodológica, as etapas do processo de investigação e os procedimentos e técnicas empregadas para atingir o objetivo do trabalho.

1.1.1. CLASSIFICAÇÃO METODOLÓGICA

A literatura apresenta diversas abordagens classificatórias para a investigação científica. Por exemplo, ao escolher uma classificação apresentada por Demo (1985), esta investigação poderia ser classificada como exclusivamente teórica, embora os resultados sejam construídos mediante as evidências de uma realidade prática, propondo uma aplicação objetiva. Isto indica que a classificação metodológica possui peculiaridades a serem consideradas e podem assumir características distintas, conforme as suas especificidades, determinados pelo tipo de objeto a investigar e pela classe de proposições a descobrir.

Para seguir uma organização clara, no propósito de melhor explicar a caracterização do trabalho, opta-se por utilizar como classificação metodológica os parâmetros abordados e referenciados por Gil (2008). Assim, este estudo será classificado metodologicamente quanto à finalidade; em relação aos níveis de investigação; ao método utilizado; ao delineamento da investigação; e, quanto a abordagem para a análise dos dados. Seguindo o autor referenciado esta investigação pode ser classificada como aplicada, pois procura responder como estruturar uma RDM (*Research Data Management*) que permita realizar uma partilha adequada dos dados, utilizando atores, articulações e práticas, através de um modelo de construção teórica útil em uma área e um local específico – a área da RDM das Unidades de Investigação e Desenvolvimento (UI&D) na Universidade de Coimbra (UC). Ou seja, a preocupação final está orientada para uma aplicação imediata em uma realidade prática, tendo como característica o interesse no uso e no conhecimento aplicado.

Quanto ao nível de investigação o caráter exploratório é adequado ao estudo, pois os levantamentos bibliográficos realizados nas principais bases de dados não indicaram resultados de investigações anteriores relacionados com os problemas abordados ou com a questão desta tese (Apolinário, 2007). Isto quer dizer que não foram identificados outros trabalhos que estivessem preocupados, especificamente e de modo conjunto, com a partilha adequada dos dados de investigação no ambiente da universidade, tendo por base a articulação dos atores e as práticas para a RDM. Cooper e Schindler (2016) e Gil (2008), afirmam que estudos exploratórios tendem a gerar estruturas com o objetivo de descobrir futuras tarefas de investigação e que desenvolvem, esclarecem e modificam ideias para a formulação de problemas e hipóteses para estudos posteriores. É também neste intento que o trabalho investiga uma área com combinação de conceitos e características abrangentes, que podem ser o ponto de partida para novos estudos na área.

Em relação ao método, o que mais este trabalho se aproxima é o do dedutivo, no sentido de que parte do geral ao particular. Ou seja, com base em premissas, segue de uma abordagem sobre os conceitos trabalhados na área e constrói-se um modelo para a aplicação no âmbito da RDM, especificamente para as UI&D da universidade. Quanto ao delineamento do trabalho, a presente investigação é considerada de cunho bibliográfico e estudo de campo, pois seus conceitos foram primeiramente levantados na literatura já

publicada para então estudar um grupo em termos de estrutura, ressaltando a interação de seus componentes e na procura de descrever com precisão as suas características. Por sua vez, em relação à abordagem para a análise dos dados, a investigação é de natureza qualitativa, pois o material obtido na bibliografia e os dados coletados através das entrevistas foram analisados e interpretados de modo textual, sem o auxílio de quantidades ou estatística para a conclusão dos resultados.

1.1.2. ETAPAS DO PROCESSO DE INVESTIGAÇÃO

O processo de investigação foi organizado em cinco etapas, quais sejam: 1) o estabelecimento da fundamentação teórica; 2) a identificação das amostras do estudo; 3) a coleta dos dados; 4) a análise da literatura e dos dados; e, 5) o relato dos resultados.

A primeira etapa – *o estabelecimento da fundamentação teórica* – foi definida em quatro vertentes, a saber: 1) as pesquisas em documentos científicos para delinear as referências conceituais atreladas à RDM e aos dados de investigação; 2) a identificação e análise via pesquisa bibliográfica de casos documentados em universidades nacionais e internacionais sobre as articulações entre atores da RDM na universidade; 3) o levantamento do material bibliográfico relativos à ciência aberta e à teoria da atividade, sendo levantada somente a parte da teoria eleita para articular com o estudo; e, 4) a pesquisa bibliográfica e o levantamento do material sobre modelos de ciclos de vida que representam os fluxos da RDM, curadoria e dos dados de investigação. Todas as vertentes desta primeira etapa serviram como parâmetros para a construção do modelo proposto e, em essência, foram utilizados para realizar a análise da literatura e contrastar com os dados empíricos.

A segunda etapa – *a identificação da amostra de estudo* – parte da premissa que a população inclui na UC, todos os centros, laboratórios, institutos e unidades envolvidas com a investigação e a gestão dos dados de investigação. Nesse entendimento, o universo de estudo abrange os investigadores, profissionais das áreas da informação, gestão, administração, órgãos, departamentos, projetos e outros atores que interagem com a comunidade científica e com a RDM. Porém, de modo a organizar e tornar viável esta investigação, foi necessário estabelecer uma identificação inicial da amostra – uma parcela convenientemente selecionada do universo da população para que o estudo seja realizado

(Marconi & Lakatos, 2003). Neste intento e para a obtenção de dados relevantes, considera-se como ambiente representativo inicial, as UI&D associadas ao Instituto de Investigação Interdisciplinar (III) da UC, que reúne 38 UI&D financiadas e avaliadas pela Fundação para a Ciência e Tecnologia (FCT). No mesmo sentido, os investigadores das UI&D são caracterizados como o subconjunto de indivíduos da amostra do estudo.

A terceira etapa – *a coleta dos dados* – foi realizada com a utilização da técnica de entrevista, recolhido em uma amostra intencional e selecionada de acordo com os interesses e a conveniência da investigação. Assim, após a identificação da amostra, a recolha dos dados se deu com investigadores pertencentes as UI&D da UC e, à medida que os dados foram coletados, surgiram outras amostras que direcionaram a novas recolhas. Portanto, a coleta foi realizada progressivamente, tendo como fio condutor uma amostra inicial e as entrevistas com os investigadores das UI&D.

O método aqui proposto tem na quarta etapa – *a análise da literatura e dos dados coletados* – o trabalho de compreensão dos dados recolhidos e da sua fundamentação. Para tal, foi sistematizado o construto teórico e os dados coletados que interessam à investigação. A sistematização do construto teórico foi realizado sob o nexos dos três eixos de estudo desta tese: 1) a articulação dos atores; 2) as práticas para a RDM; e, 3) a partilha dos dados de investigação. E o trabalho de sistematização dos dados coletados foram orientados sob três categorias propostas: 1) contexto; 2) atores; e, 3) práticas. As categorias também se desdobraram em subcategorias com seus respectivos indicadores (Tabela 3). Assim, através de instrumentos analíticos manuais, a análise dos dados e da literatura foram sistematizados para proporcionar a apresentação dos resultados.

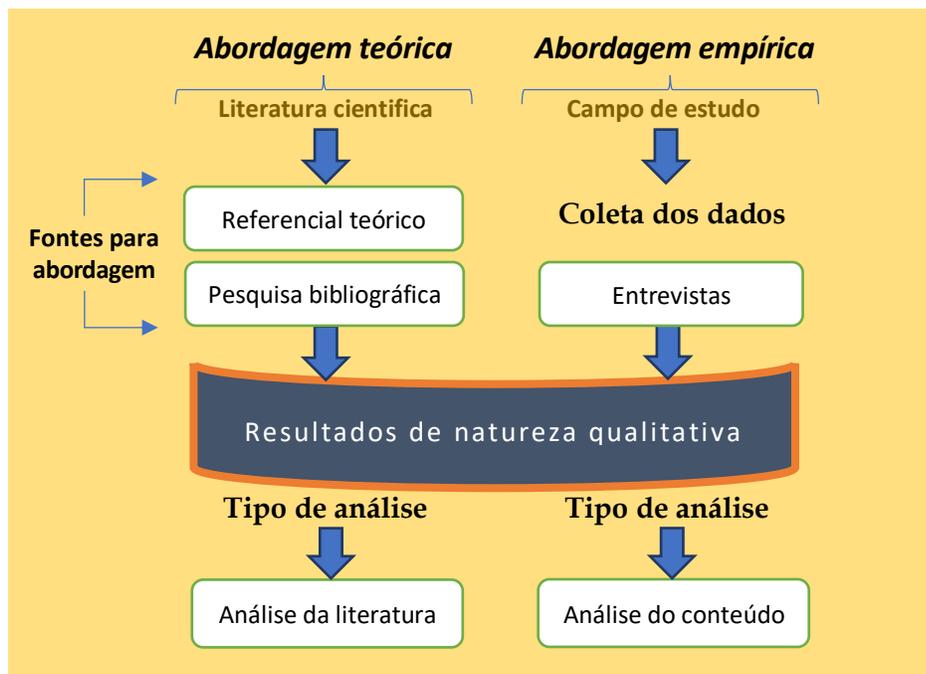
Por fim, na quinta etapa têm-se o relato dos resultados divididos em dois procedimentos, a saber: 1) a apresentação e interpretação de modo sintetizado, dos dados coletados e da análise da literatura, articulando a teoria com a realidade prática; e, 2) a apresentação e a interpretação do modelo resultante, obtido à luz da questão levantada e dos objetivos do estudo, de modo a relacionar os resultados alcançados, salientar as limitações, os aspectos pertinentes e as perspectivas futuras com o que já foi apresentado de forma estruturada no modelo.

1.1.3. PROCEDIMENTOS E TÉCNICAS EMPREGADAS

A investigação científica depende de um conjunto de procedimentos intelectuais e técnicos para que seus objetivos sejam atingidos (Gil, 2008). Os procedimentos técnicos empregados para a fundamentação e análise da teoria desta investigação, bem como para a obtenção e a análise dos dados do campo empírico, encaminharam-se sob as respectivas abordagens apresentadas na figura 1. Na abordagem teórica foram usadas duas técnicas: a concepção do referencial teórico e a pesquisa bibliográfica. Para a abordagem empírica, usou-se a coleta dos dados através da aplicação de entrevistas no campo de estudo. Os resultados, de natureza qualitativa, foram examinados tendo por base a técnica da análise do conteúdo e a análise da literatura. O construto teórico e o detalhamento dos procedimentos e técnicas para as referidas abordagens, estão especificados a seguir.

Figura 1:

Procedimentos e técnicas empregadas



1.1.3.1. ABORDAGEM TEÓRICA

Para abordagem e formulação do referencial teórico, Demo (1985) descreve três procedimentos. O primeiro diz respeito ao domínio da literatura da área temática, que carrega a acumulação já feita de conhecimento, os conflitos das correntes de pensamentos,

a cristalização de práticas de investigação e o ambiente da discussão atual acerca do assunto. O segundo procedimento é o domínio da literatura fundamental, através da qual se toma conhecimento da produção existente, estimulando a criatividade e o diálogo crítico. Assim, construções teóricas novas e originais têm sua gênese também em construtos estabelecidos anteriormente, a partir de resultados alcançados por outros investigadores. O terceiro procedimento, decorrente dos dois anteriores, está na vivacidade crítica, na qual se instala a discussão aberta como meio para o crescimento científico.

Sob os procedimentos supracitados, o referencial teórico foi concebido de modo que o modelo proposto no estudo representasse a relação entre as articulações dos atores, as práticas para gestão e a partilha dos dados de investigação, no ambiente da universidade. Esta relação é suportada respectivamente, em termos teóricos e conceituais, pela teoria da atividade, por modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação e pela ciência aberta. A contextualização do ambiente temático da RDM na universidade foi outra atividade decorrente da fundamentação teórica.

O material bibliográfico foi pesquisado e identificado principalmente através de portais integrados de pesquisa. A seleção do material das pesquisas, que se soma a literatura já conhecida da área e à literatura fundamental, foi realizada sob dois propósitos: a contextualização da temática relacionada a RDM e a universidade e, a obtenção dos modelos de ciclos de vida, que representam os fluxos da RDM, da curadoria e dos dados de investigação. Por se tratar de uma área do conhecimento interdisciplinar, parte dos fundamentos que subsidiam o estudo advém das disciplinas com as quais a ciência da informação mantém vinculação conceitual e metodológica.

As fontes das pesquisas foram a *b-on* – Biblioteca do Conhecimento *Online*⁴, a *Web of Science* e a *Library & Information Science Source*. De modo específico, para obter o material sobre a relação entre a RDM e a universidade, usaram-se os termos “*research data management*”, “*university*”, “*college*” e “*higher education*” combinados por operadores booleanos na seguinte frase de pesquisa (*query*): “*research data management*” AND “*university*” OR “*college*” OR “*higher education*”. No mesmo sentido, para a

⁴ A *b-on* reúne as principais editoras de revistas científicas internacionais de modo a oferecer acesso a milhares de periódicos e artigos científicos em texto integral - <https://www.b-on.pt/>.

identificação do material dedicado aos modelos de ciclos de vida, os termos da pesquisa incluíram qualquer combinação das três seguintes palavras: “*research*”, “*data*” e “*lifecycle*”, os resultados foram somados a uma lista de artigos já conhecidos.

O resultado da pesquisa bibliográfica e do referencial teórico fundamentou a estrutura conceitual que norteia o desenvolvimento da investigação e, ao mesmo tempo, construiu a relação conceitual entre a articulação dos atores, as práticas para a RDM e a partilha dos dados de investigação. Deste modo, a análise de obras relevantes acerca dos temas: ciência aberta, teoria da atividade e os modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, resultaram nas sintetizações realizadas no decorrer do trabalho, apresentadas e discutidas no capítulo 6.

Ao associar os objetivos específicos desta tese, com os procedimentos e as técnicas empregadas pela abordagem teórica, verifica-se que estes serviram de forma direta para a realização de cinco objetivos do estudo: 1) contextualizar o ambiente temático da RDM na universidade; 2) caracterizar os dados de investigação e identificar os atores envolvidos na relação entre a RDM e a universidade; 3) obter os parâmetros necessários para definir uma partilha adequada dos dados de investigação; 4) identificar as práticas essenciais para a RDM na partilha dos dados; e, 5) propor uma estrutura para articulação dos atores da RDM. O resultado destes objetivos também estabelece o nexos teórico entre os eixos do estudo. A associação dos objetivos com a área temática e o procedimento, estão resumidos a seguir.

Tabela 1:

Objetivos da abordagem teórica

Objetivos para o estudo	Área temática	Procedimento
1. Contextualizar o ambiente temático da RDM na universidade.	RDM e universidade	Pesquisa bibliográfica
2. Caracterizar os dados de investigação e identificar os atores do ambiente temático.	RDM e universidade	Levantamento direcionado na literatura da área temática
3. Parametrizar para definir uma partilha adequada dos dados de investigação.	Ciência Aberta	Levantamento direcionado na literatura fundamental
4. Identificar as práticas essenciais para a RDM na partilha dos dados.	Modelos de ciclos de vida da RDM, curadoria e dos dados	Pesquisa bibliográfica e Levantamento direcionado na literatura da área temática
5. Propor uma forma para articular os atores com a finalidade de gerir e partilhar os dados.	Teoria da Atividade	Levantamento direcionado na literatura fundamental

1.1.3.2. ABORDAGEM EMPÍRICA

A atividade da coleta dos dados ocorreu através da aplicação de entrevistas. Esta técnica foi considerada apropriada, pela finalidade exploratória do estudo e de modo a obter o maior número possível de informações sobre o tema (Minayo, 2001), tendo em vista que de acordo com Quivy e Campenhoudt (1988) o entrevistado é levado, com frequência, a refletir sobre o assunto analisado, fazendo com que a comunicação resultante seja um processo de elaboração de um pensamento e não apenas de um dado coletado.

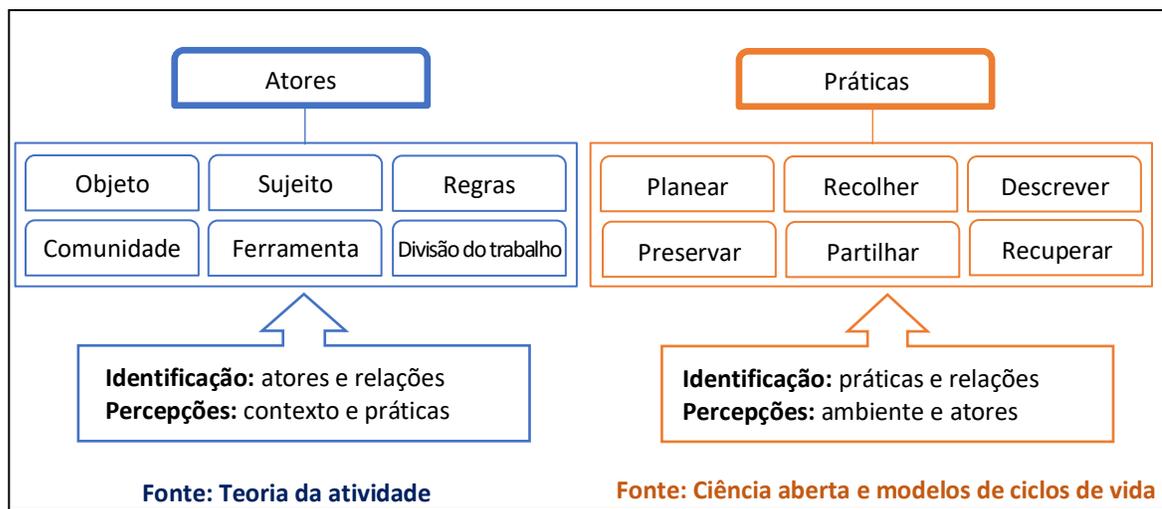
Ao considerar o ambiente disciplinar diverso e as especificidades das investigações, fez-se por bem adotar a entrevista semiestruturada – conveniente pela flexibilidade em relação à ordem da realização das perguntas –. A entrevista permitiu obter respostas mais próximas da linguagem do sujeito, que aproximaram mais da compreensão da realidade dos elementos em estudo (Quivy & Campenhoudt, 1998; Valles, 2007).

A categorização para a coleta e análise dos dados, foi definida de modo prévio, de acordo com o construto teórico e os objetivos da tese. Nesse sentido, Bardin (2011) indica a possibilidade de uma categorização a priori, sugerida pelo referencial teórico. Seguindo a mesma direção Oliveira et al. (2003), afirmam que na construção de categorias a priori, deve-se levar em consideração a orientação teórica e os objetivos da investigação. Estes últimos autores ainda afirmam que a categorização pode variar de uma generalidade mais forte até uma generalidade mais fraca, caracterizando as subcategorias aqui propostas.

Ao abrigo das referências acima, os blocos temáticos, bem como as categorias, subcategorias e as respectivas finalidades (Tabela 2), foram elaboradas orientados para a realização de uma RDM com a finalidade da partilha dos dados, que articulasse atores e realizasse práticas para o intento, no ambiente da universidade, ou seja, tiveram por base para a sua construção o objetivo desta tese. Assim, identificar e caracterizar tais elementos em contexto, juntamente com os relacionamentos e as percepções que os rege (Figura 2), foram os aspectos norteadores para a categorização analítica proposta. O resultado foi o agrupamento dos elementos teóricos com características comuns, ou relacionados entre si, organizados de modo a possibilitar a realização das entrevistas e a consequente coleta dos dados que interessam a este estudo.

Figura 2

Elementos teóricos para a construção das categorias analíticas



A organização, adaptação e agrupamento dos elementos teóricos da figura acima em blocos, com suas respectivas categorias, subcategorias e finalidades temáticas, resultam na categorização para a coleta e análise dos dados, conforme a tabela abaixo.

Tabela 2

Categorização para a coleta e análise dos dados.

Blocos temáticos	Categorias	Subcategorias	Finalidade
I. Ambiente e características dos dados de investigação.	Contexto	Ambiente	Perceber o contexto no qual os dados de investigação estão sendo administrados, assim como obter informações em relação as características e a utilidade dos dados geridos no ambiente da investigação.
		Dados	
II. Recursos para a RDM.	Atores	Pessoas	Obter informações sobre os atores e as articulações para as práticas da RDM, através da identificação das pessoas responsáveis, dos processos e dos fluxos de trabalho existentes, dos colaboradores e das colaborações que influenciam na atividade, assim como as ferramentas e serviços, as políticas, normas ou orientações adotadas no apoio da RDM para o contexto.
		Fluxos de trabalho	
		Colaborações	
		Ferramentas	
		Políticas	
III. Hábitos e percepções para a partilha dos dados em acesso aberto.	Práticas	Partilha	Verificar as práticas e obter as percepções em relação aos fatores que habilitam ou restringem a partilha dos dados de investigação em acesso aberto.
		Acesso aberto	

Para obter um melhor entendimento do que se deseja conhecer sobre o contexto, as práticas, os atores e suas articulações, a tabela 3 apresenta os indicadores e objetivos específicos das subcategorias, de modo a detalhar a categorização proposta anteriormente.

Tabela 3

Detalhamento da categorização para a coleta e análise dos dados

Categories	Subcategorias	Indicadores	Objetivos específico
1. Contexto	Bloco temático I: Ambiente e características dos dados de investigação		
	1.1. Ambiente	<ul style="list-style-type: none"> • Domínio Científico • Propósito da investigação • Constituição da equipe • Modo de condução e capacidade operacional 	Constatar a finalidade, o escopo disciplinar e as peculiaridades onde a investigação é realizada.
	1.2..Dados	<ul style="list-style-type: none"> • Características • Utilidade interdisciplinar • Utilidade social • Sensibilidade 	Conhecer a fonte, a natureza, a materialidade, o valor e a utilidade dos dados geridos.
2. Atores	Bloco temático II: Recursos para gestão dos dados de investigação		
	2.1. Pessoas	<ul style="list-style-type: none"> • Responsáveis e funções • Envolvimento na atividade • Conhecimentos sobre o assunto 	Identificar as pessoas envolvidas de forma direta ou indireta na atividade.
	2.2. Fluxos de trabalho	<ul style="list-style-type: none"> • Divisão formal ou informal das atribuições • Processos estabelecidos • Atividades segmentadas 	Constatar as atribuições e fluxos na organização processual do trabalho.
	2.3. Colaborações	<ul style="list-style-type: none"> • Parcerias com instituições • Vínculos com indivíduos e profissionais • Possíveis cooperações 	Averiguar a influência na atividade por profissionais e instituições internas ou externas à universidade.
	2.4. Ferramentas	<ul style="list-style-type: none"> • Armazenamento • Segurança • Serviços e gestão • Meios de partilha 	Verificar a utilização da estrutura técnica e de serviços para a RDM.
	2.5. Políticas	<ul style="list-style-type: none"> • Conhecimento e utilização • Elaboração e aplicação • Plano de gestão dos dados e metadados 	Perceber o uso de normas específicas ou políticas gerais para a RDM.
3. Práticas	Bloco temático III: Hábitos e percepções para a partilha dos dados em acesso aberto		
	3.1. Partilha	<ul style="list-style-type: none"> • Necessidades • Formas • Dificuldades 	Compreender as demandas, as práticas e as barreiras para a partilha.
	3.2. Acesso aberto	<ul style="list-style-type: none"> • Interesse • Preocupações • Motivações 	Reconhecer os aspectos que influenciam as práticas para a abertura dos dados de investigação.

As entrevistas aplicadas sob a categorização apresentada, cujo documento de convite e o roteiro contendo as perguntas estão nos apêndices da tese, foram conduzidas de forma *online*, em dia e hora previamente agendados, com investigadores das UI&D da UC. O grupo de investigadores inicialmente convidado, foi obtido por intermédio do III da UC, que elaborou e entregou em 31 de agosto de 2020 uma lista com o nome das UI&D, as possíveis características a serem exploradas em relação a RDM e o contato com o responsável pela unidade de investigação. Tendo em vista o interesse na observação da diversidade, foram convidados investigadores pertencentes aos seis domínios científicos definidos de acordo com as categorias do III (Tabela 5).

No intento de obter mais respondentes, os entrevistados também indicaram outros investigadores e projetos que pudessem colaborar com a temática das perguntas. Desta forma, chegou-se a um total de vinte e um investigadores selecionados e convidados para as entrevistas. Destes, responderam positivamente ao convite onze investigadores com dados geridos em cinco domínios científicos diferentes. As entrevistas ocorreram entre 16 de outubro de 2020 até 31 de janeiro de 2022. Cabe destacar que os investigadores que aceitaram o convite têm perfis variados, também exercendo funções diretivas e acadêmicas.

As perguntas foram elaboradas seguindo a categorização proposta, com a finalidade de compreender a realidade dos elementos em estudo e, como já referenciado, para a comparação da teoria com a realidade. As respostas tiveram a gravação em áudio e vídeo realizadas conforme o termo de consentimento. Embora existissem perguntas sobre a identificação do projeto e do respondente, nenhum nome ou informação que identificasse as pessoas foram expostos ou utilizados, senão para fins de administração interna dos dados.

Em relação aos procedimentos para interpretação dos dados coletados, a metodologia utilizada foi a análise de conteúdo, que segundo Bardin (2011) é caracterizada por um conjunto de técnicas de análise das comunicações, tendo por objetivo realizar a descrição do conteúdo das mensagens, permitindo a inferência de conhecimentos relativos às condições de produção ou recepção destas mensagens. Assim, seguindo a metodologia proposta pela autora, o material foi trabalhado no tratamento dos resultados, pelo qual foram inferidos e interpretados os dados brutos, ou seja, as respostas das entrevistas, agrupando-os de acordo com a categorização proposta para posterior análise e discussão.

A análise e discussão dos dados obtidos por meio dos levantamentos da teoria e no campo empírico, são apresentados de modo integrado no capítulo 6. Primeiro apresenta-se o resultado da consolidação teórica e depois os dados coletados por meios das entrevistas. Ou seja, os dados coletados e o construto teórico foram sintetizados, analisados e discutidos para obter o modelo proposto. Portanto, o resultado da análise do material coletado e da discussão pertinente no confronto da teoria com a realidade, obteve os parâmetros pertinentes à construção do modelo para a RDM, finalidade principal deste estudo. A tabela 4 resume as técnicas e procedimentos apresentados nesta seção, de modo a ser visualizado: a relação entre os objetivos específicos da tese; as fontes de coleta dos dados; os instrumentos utilizados para atingir o objetivo específico, além do principal tipo de análise realizada.

Tabela 4

Relação entre os objetivos específicos, as fontes, os instrumentos e a análise dos dados

Objetivos específicos	Fontes	Instrumentos	Análise
1. Contextualizar, fundamentado na literatura, o ambiente temático da RDM na universidade, identificando os atores do contexto e os dados de investigação.	Base de dados da literatura científica	Pesquisa bibliográfica Referencial teórico	Análise do conteúdo
2. Caracterizar uma partilha adequada dos dados de investigação, através dos fundamentos da ciência aberta.	Base de dados da literatura científica	Referencial teórico	Análise da literatura
3. Identificar as práticas que sejam essenciais para a gestão e partilha dos dados, apoiado por modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, incluindo seus elementos e processos.	Base de dados de literatura científica	Pesquisa bibliográfica Referencial teórico	Análise de conteúdo
4. Estabelecer uma proposta para a articulação dos atores e a adaptação contextual para a RDM, viabilizado pelos fundamentos da teoria da atividade.	Base de dados da literatura científica	Referencial teórico	Análise da literatura
5. Propor umnexo teórico entre a articulação dos atores, as práticas essenciais para a RDM e a partilha adequada dos dados de investigação.	Base de dados da literatura científica	Referencial teórico	Análise da literatura
6. Perceber, com base na compreensão de investigadores de UI&D da UC, o ambiente e os dados de investigação para a RDM, o relacionamento dos atores e as práticas envolvidas, bem como os hábitos para a partilha dos dados.	Investigadores	Entrevistas	Análise do conteúdo

1.2. CAMPO EMPÍRICO – A UNIVERSIDADE DE COIMBRA

O campo empírico encontra-se sob o ambiente organizacional da UC – uma instituição com mais de sete séculos de história, constituída por oito faculdades, sendo que em 31 de dezembro de 2021, contava com 2.126 docentes e investigadores, distribuídos em 42 unidades e 483 projetos de investigação, com uma população estudantil de mais de 25 mil alunos (Universidade de Coimbra, 2022). Conforme explicado na seção anterior, por ser considerado inviável o estudo em todo o universo da UC, definiu-se como ambiente representativo adequado para o estudo as 38 UI&D associadas aos III, financiadas e avaliadas pela FCT e distribuídas em 6 áreas científicas, conforme detalhadas na tabela que segue.

Tabela 5

Relação das 38 UI&D da Universidade de Coimbra e os domínios científicos

Domínio científico	Unidades de Investigação
1. Artes e Humanidades	Centro de Estudos Interdisciplinares do Século XX (CEIS20)
	Centro de Estudos em Arqueologia, Artes e Ciências do Patrimônio (CEAACP)
	Centro de História da Sociedade e da Cultura (CHSC)
	Centro de Estudos Clássicos e Humanísticos da Faculdade de Letras da Universidade de Coimbra (CHEC)
	Centro Interuniversitário de Estudos Camonianos (CIEC)
	Instituto de Estudos Filosóficos (IEF)
2. Ciências da Saúde	Centro de Neurociências e Biologia Celular (CNC)
	Centro de Imagem Biomédica e Investigação Translacional (CIBIT)
	Instituto de Investigação Clínica e Biomédica de Coimbra (ICBR)
	Centro de Investigação do Desporto e da Atividade Física (CDAF)
	Centro de Inovação em Biomedicina e Biotecnologia (CIBB)
Centro de Estudos e Investigação em Saúde (CEISUC)	
3. Ciências Exatas	Centro de Física da Universidade de Coimbra (CFisUC)
	Laboratório de Instrumentação e Física Experimental de Partículas (LIP)
	Unidade de Investigação e Desenvolvimento Química-Física Molecular
	Centro de Matemática da Universidade de Coimbra (CMUC)
	Centro de Química de Coimbra (CQ)
Laboratório de Instrumentação, Engenharia Biomédica e Física das Radiações (LIBPhys)	
4. Ciências Naturais e do Ambiente	Centro de Ciências do Mar e do Ambiente (MARE)
	Centro de Geociências (CGEO)
	Centro de Investigação da Terra e do Espaço da Universidade de Coimbra (CITEUC)
	Centro de Ecologia Funcional (CFE)
5. Ciências Sociais	Centro de Estudos Sociais (CES)
	Instituto Jurídico da Faculdade de Direito da Universidade de Coimbra
	Centro de Investigação em Antropologia e Saúde (CIAS)
	Centro em Rede de Investigação em Antropologia (CRIA)

	Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental (CINEICC)
	Centro de Estudos de Geografia e Ordenamento do Território (CEGOT)
	Instituto de Psicologia Cognitiva, Desenvolvimento Humano e Social
	Centro de Investigação em Economia e Gestão da Universidade de Coimbra (CEBER)
6. Engenharia e Tecnologias	Centro de Informática e Sistemas da Universidade de Coimbra (CISUC)
	Instituto de Engenharia de Sistemas e Computadores de Coimbra (INESC)
	Centro de Investigação em Engenharia dos Processos Químicos e dos Produtos da Floresta (CIEPQPF)
	Centro de Engenharia Mecânica, Materiais e Processo (CEMMPRE)
	Centro de Investigação do Território, Transporte e do Ambiente (CITTA)
	Instituto para a Sustentabilidade e Inovação em Estruturas de Engenharia (ISISE)
	Instituto de Sistemas e Robótica (ISR)
	Instituto de Telecomunicações (IT)
	Centro de Investigação em Engenharia dos Processos Químicos e dos Produtos da Floresta (CIEPQPF)
	Laboratório Associado de Energia, Transportes e Aeronáutica (LAETA)

Nota. Obtido em: <https://ucpages.uc.pt/iii/investigacao/unidades-i-d-e-laboratorios-associados/>

As UI&D, de modo geral, apresentam sua importância no fato de que a maior parte da investigação científica portuguesa é ali realizada, juntamente com laboratórios associados, sendo alvos de avaliação e financiamento pela FCT. É importante destacar no conhecimento do campo empírico, que no ambiente da atividade científica e tecnológica em Portugal, a FCT é a instância do governo que tem sob sua responsabilidade a incumbência de se relacionar direta e indiretamente com as UI&D. Neste sentido, a FCT reconhece a importância das UI&D ao afirmar ser este um pilar fundamental na consolidação do sistema científico moderno e competitivo, como base da organização científica e tecnológica do país⁵. Sob tal contexto, a FCT tem influência na promoção do desenvolvimento das UI&D, sendo responsável por aperfeiçoar o ambiente institucional e legal, ampliar as fontes e volumes de recursos, estabelecer instrumentos, mecanismos e programas adequados às necessidades de consolidação das unidades, articulando e fortalecendo seus atores e atividades científicas, bem como as atividades tecnológicas inovadoras (Fundação para a Ciência e Tecnologia [FCT], 2021).

As UI&D, por sua vez, são instituições de investigação sem fins lucrativos, que se dedicam à investigação científica e ao desenvolvimento tecnológico, é o ambiente onde os investigadores realizam os projetos científicos. Neste âmbito, assumem particular

⁵ Tendo por base a “lei da ciência” - Decreto-Lei nº 63/2019 - <https://dre.pt/application/conteudo/122317422>

importância na promoção e no reforço de competências das instituições científicas e tecnológicas, através da participação das suas equipas em projetos de investigação científica e desenvolvimento tecnológico por diferentes domínios científicos. No entanto, as UI&D têm as atividades circunscritas em áreas disciplinares específicas, que em seu escopo de atividade, é orientada a “reunir recursos interdisciplinares e multidisciplinares que potenciem a abordagem de problemas complexos e novos desafios sociais” (FCT, 2021, para. 1).

Embora possam estar ligadas ao ensino, as atividades das UI&D são direcionadas à produção do conhecimento científico, por meio da investigação, desenvolvimento e inovação. Pelo fato de estarem em áreas do conhecimento específicas e por terem suas finalidades dirigidas para atividades de produção do conhecimento científico, assume-se que tais instituições constituem ambientes férteis para a realização desta investigação. Nas universidades, as UI&D podem funcionar como entidades integradas ou parceiras, ou ainda possuírem um estatuto independente e funcionamento autónomo. Cabe enfatizar que no ambiente das UI&D, os investigadores são os principais atores e usuários essenciais no processo para a gestão dos dados de investigação.

Especificamente em relação ao ambiente das UI&D da Universidade de Coimbra, o relatório da comissão de investigação do conselho geral da UC⁶ (Relatório UC, 2019), somado ao documento para a organização, gestão e regulação das atividades de investigação da UC⁷, apresentam um panorama relevante do ambiente na UC, em relação aos recursos humanos, financeiros, espaços físicos e produção científica. Assim, destaca-se a seguir algumas constatações extraídas destes documentos, que ajudam a compreender o ambiente das UI&D como campo empírico e direcionados aos objetivos discutidos nesta tese.

- O financiamento das UI&D parte de duas fontes distintas: a UC e as próprias unidades;
- Se elege a interdisciplinaridade como uma vertente de trabalho importante;

⁶ https://www.uc.pt/governo/cons_geral/docs_aprovados_eventos/doc_CI

⁷ Embora este documento não esteja vigente, os parâmetros extraídos são úteis para compreensão das UI&D, tal documento foi recuperado em: https://www.uc.pt/regulamentos/transversais/nao-vigentes/regulacao_atividades_investigacao_cientifica_modalidade_custos_totais_desenvolvidas_uc_parte2.pdf

- Distingue-se as UI&D entre si pela natureza jurídica: ou são unidades integradas na UC ou são unidades autônomas;
- A maioria das UI&D ligadas à UC desenvolve as atividades dentro da UC;
- As UI&D têm captado financiamentos predominantemente de natureza adicional, principalmente da FCT ou através de projetos europeus;
- Independentemente do financiamento, existem sempre custos de estrutura, necessários ao suporte geral às atividades e dos projetos, os custos são financiados pela própria UC em parte significativa;
- a UC está presente nas seis grandes áreas do conhecimento definidas pela FCT, com um número relevante de unidades e de investigadores em cada uma dessas áreas, e sem que se possa identificar uma área predominante;
- Há UI&D de dimensão relativamente grande que concentram muitos dos investigadores de uma área;
- Os investigadores doutorados são maioritariamente docentes na UC;
- As 38 Unidades de UI&D incluem um total de 14.8% do total de investigadores doutorados em Portugal;
- Articulação entre as unidades de investigação, carece de sistemas de informação que respondam a necessidades de informação de docentes e investigadores, que procuram potenciais colaborações e valências;
- Existe carência de apoio na maioria das UI&D integradas na UC e mesmo nas unidades autônomas;
- As UI&D autônomas têm menos apoio dos serviços da UC, o que frequentemente leva à criação de um quadro de funcionários próprio;
- As UI&D autônomas têm por vezes de pagar pelas instalações/equipamentos que usam;
- As UI&D autônomas tem sentido por vezes que não é reconhecido o seu importante papel para a produção e o prestígio científico da UC;
- Recomenda-se a articulação dos seus centros de investigação, promovendo a melhoria da qualidade das estruturas de investigação e a motivação generalizada;

- As UI&D prosseguem sobretudo o seu interesse próprio, muitas vezes em competição com outras UI&D do mesmo Grupo UC, cujas capacidades até podem desconhecer ou ignorar;
- Existem poucos fóruns de articulação, a UC poderia ser mais competitiva nacional e internacionalmente se esta vertente fosse potenciada, era importante reunir “boas práticas”, e ajudar a torná-las mais transversais a toda a UC;
- A vertente de investigação não deve ser pensada desligada da dimensão pedagógica inerente à universidade;
- A colaboração com parceiros externos para produzir resultados económicos e sociais que extravasam a produção académica tradicional define o chamado impacto da investigação; e
- A ligação à sociedade, nos seus mais diversos aspectos, é uma das missões aglutinadoras da universidade.

É no sentido do ambiente percebido acima, que este trabalho encontra nas UI&D da UC, o campo empírico julgado como ideal para o estudo, pois na diversidade das áreas de atuação e resguardadas as diferenças essenciais entre elas, as atividades de investigação viabilizadas pelos investigadores estão relacionadas, nos procedimentos, nos hábitos e nos padrões, tanto com os pressupostos científicos, quanto com a RDM e seus atores, mas também com as articulações e as práticas para a partilha dos dados. Isto porque, embora distintas em sua natureza disciplinar, as atividades de investigação realizadas pelas UI&D guardam princípios científicos comuns para a produção dos dados, pautados, evidentemente, pelo método científico. Assim, apesar de diferentes, a noção que converge com a finalidade de todas as 38 unidades de investigação, bem como da própria universidade, é a elaboração e a partilha universal do conhecimento.

2. A GESTÃO DOS DADOS DE INVESTIGAÇÃO NA UNIVERSIDADE

A abordagem deste capítulo apresenta uma contextualização teórica sobre o universo do estudo – a RDM (*Research Data Management*) na universidade. Tal contextualização é orientada pela visão relacional do ambiente temático e tem triplo objetivo, qual seja: a) exibir um panorama do ambiente da RDM na universidade; b) identificar atores⁸ participantes do ambiente; e, c) caracterizar o objeto informacional a ser gerido – os dados de investigação. Em conjunto, estes objetivos apresentam elementos teóricos que ajudam no desenvolvimento deste trabalho e na resposta da questão complementar: quem são os atores da RDM na universidade?

Para alcançar a finalidade declarada, o presente capítulo foi organizado em quatro seções. Inicia-se com um referencial teórico sobre a relação entre a RDM e a universidade, depois através de pesquisa bibliográfica, é abordada a relação dos atores na articulação com a RDM e com o ambiente e, ainda, é referenciada a relação entre os dados de investigação com a RDM na universidade. Na última parte, para um melhor entendimento, realiza-se uma síntese do capítulo. Sob esta abordagem, cumpre-se o primeiro objetivo específico da tese.

2.1. A RELAÇÃO NO AMBIENTE TEMÁTICO

A relação entre a RDM e a universidade parte do pressuposto da necessidade de tal vínculo, presumindo que as universidades precisam enfrentar os desafios advindos da produção de grandes quantidades de dados de investigação, sob a importância com que eles são tratados na atualidade e pelas conseqüentes transformações na qual passa a investigação científica. Como referenciado na introdução deste trabalho e ratificado a seguir, uma gestão adequada dos dados de investigação propicia ganhos institucionais, colaboração com a eficiência da investigação e atendimento às exigências internas e externas oriundas de vários quadrantes, estes são exemplos de fatores específicos que fazem com que as universidades necessitem incorporar práticas de RDM em seu ambiente institucional.

⁸ O termo “ator”, a ser utilizado em toda esta tese, é considerado algo que desempenhe um ato e modifique um status, seja o ator humano ou não humano. Neste sentido, atores humanos e não humanos podem se associar, interagir, combinar ou trocar propriedades entre si, assumindo novas formas e dando ao contexto um novo sentido para a realização de dada tarefa social (Callon, 1986; Latour, 1988, 2012; Law, 1987).

A necessidade e a decorrente importância da RDM na universidade, encontram fundamentos nos trabalhos de Borgman (2010, 2018) e Katz et al. (2018), quando os autores descrevem que as universidades, como produtoras e administradoras de grandes quantidades de dados, estão diante de novas oportunidades de investigação, ensino, administração e parcerias e que, a eficiência na exploração e na gestão dos dados de investigação significa tornar as universidades mais competitivas, com maior probabilidade de vantagens em longo prazo, em forma de subsídios, prêmios, melhores estudantes e recompensas.

Diante do contexto e do conjunto de possibilidades apresentadas, Borgman (2012) afirma que as: “Universities are assessing their rights, roles, and responsibilities for managing and for exploiting data from their researchers” (p.3). Shearer (2015) acrescenta que um dos fatores que contribui para o uso apropriado da RDM no ambiente das universidades é o financiamento da ciência. Tal afirmação encontra fundamento na declaração da Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2021), assinada em 2006 por 34 países e atualizada em 2021, sobre a necessidade de acesso aos dados de investigação provenientes de financiamento público. Mas também é observada nas iniciativas de agências financiadoras da ciência em todo mundo, que passam a adotar e exigir como requisito fundamental para investimento, a prática de políticas referentes à RDM para os dados produzidos no âmbito das investigações.

Simultaneamente à observação de que a realização da RDM é fundamental para as universidades, sua aplicação não é simples e está sob diversos desafios, grande parte provenientes da heterogeneidade disciplinar, humana e tecnológica do ambiente. Neste contexto complexo, existe uma combinação necessária de atores humanos e não humanos onde, tais atores precisam de coordenação e definição clara de responsabilidades para que não haja justaposição de ações, o que impõe desafios e formas de organização planejadas e ajustadas. Portanto, a abordagem no contexto em questão, requer o enfrentamento de desafios nos aspectos técnicos e sociológicos (e-IRG, 2016).

Os desafios sociológicos envolvem vários aspectos da condição humana, ou seja, suas características fisiológicas e psicológicas, seus níveis de qualificação, suas relações sociais dentro da instituição e as condições organizacionais do trabalho, que compreende tarefas a serem realizadas e as condições técnicas para sua realização, envolvendo assim o

ambiente de trabalho, as instalações, equipamentos, as ferramentas, os procedimentos, normas operacionais, entre outros. Nesse sentido, Shearer (2015) afirma que as instituições e os investigadores ainda precisam ser convencidos de que a RDM deva ser prioridade para a comunidade da investigação. Tal convencimento, segundo a autora, passa por resolver preocupações sobre o tempo gasto, conhecimento, transparência, proteção da privacidade, propriedade intelectual e os recursos envolvidos na preparação dos dados, bem como uma falta de experiência generalizada sobre o assunto dentro desta comunidade.

Entre os aspectos técnicos, Borgman (2018) e Sayão e Sales (2019) referenciam que poucas universidades têm mecanismos adequados de governança para enfrentar os desafios da RDM de forma eficaz, uma vez que a natureza variada dos dados exige infraestruturas e estratégias diversificadas para que sejam geridos. Some-se a isto, a preocupação das instituições sobre como financiarão os serviços de apoio a RDM. Katz et al. (2018), explicam que estas preocupações acontecem porque muitas infraestruturas e práticas de investigação ainda necessitam ser desenvolvidas, com isso as universidades mostram mais interesses e intenções do que efetivamente ações de investimento ou implementações para a RDM, resultando em uma relativa lentidão no engajamento e apenas o desenvolvimento de soluções parciais.

Não obstante os desafios sociotécnicos apresentados, segundo Shearer (2015), a situação começa a mudar. No mesmo sentido, Brown, Bruce e Kernohan (2015) reconhecem algum progresso na área. O progresso e a mudança podem ser observados em diversas ações e iniciativas práticas para a RDM, como, por exemplo, na iniciativa da *League of European Research Universities* (LERU) que elenca uma série de estudos realizados em diferentes universidades, tais como: *University of Vienna*; *University College London* (UCL); *University of the West Indies* (UWI); *University of London*; e, *University of Amsterdam* (Achard et al., 2018). Do mesmo modo, a *Online Computer Library Center* (OCLC) referencia boas práticas achadas em outras instituições, tais como: *University of Edinburgh* (UK); *University of Illinois at Urbana-Champaign* (US); *Monash University* (Australia); e, *Wageningen University & Research* (Holanda) (Bryant, Lavoie & Malpas, 2017).

É no contexto e perspectivas de ganhos, desafios e mudanças que as universidades necessitam construir um maior relacionamento entre seus atores para realizar a RDM.

Assim, além de reconhecer a importância, a necessidade, os problemas e progressos em relação a gestão dos dados, é necessário também desenvolver uma visão relacional, de modo a perceber a amplitude que a RDM pode alcançar no ambiente em questão. Nesse sentido, são várias as definições que contribuem para o entendimento da RDM e o seu ambiente. Por exemplo, o *Open Access Infrastructure for Research in Europe* (OpenAIRE, 2017) define assim a RDM: “Data management means all the processes and activities required to manage data through the research life-cycle for current and future research purposes and users” (p. 3).

De acordo com Whyte e Tedds (2011), a RDM “[...] concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information” (p.1), ou como apresentado pelo FOSTER, a RDM “[...] comprises the disciplines related to managing research data as a valuable resource. These can include, but are not limited to, research data governance, security and quality” (Facilitate Open Science Training for European Research [FOSTER], 2020, para. 1).

Cox e Pinfield (2014) também asseveram que a RDM

[...] consists of a number of different activities and processes associated with the data lifecycle, involving the design and creation of data, storage, security, preservation, retrieval, sharing, and reuse, all taking into account technical capabilities, ethical considerations, legal issues and governance frameworks (p. 2).

Diante do exposto, um caminho possível para visualizar o relacionamento que permeia a RDM em uma universidade, é extrair, obviamente sem esgotar, alguns temas ou práticas com base nessas definições da RDM, conforme segue: a) processos e atividades no ciclo de vida da investigação; b) utilização atual e futura dos dados; c) divulgação; d) arquivamento; e) organização; f) verificação dos resultados; g) investigações novas e inovadoras; h) armazenamento; i) segurança; j) preservação; k) recuperação; l) partilha; m) reutilização; n) capacidades técnicas; o) questões legais; p) estruturas de governança; e, q) qualidade dos dados. Percebe-se que a RDM aborda assuntos variados, que juntos devem ser alvo de atenção detalhada, pois são intercalados e carecem de uma organização lógica, com questões de amplitude abrangentes do tipo “estruturas de governança”, ou específicas como a “preservação dos dados”, porém ambas devem ser tratadas sob igual importância, dada a dependência de uma sobre a outra, com a RDM e com o ambiente.

Em relação à definição da RDM e a sua abrangência, neste trabalho entende-se que as múltiplas características da RDM não permitem adotar uma definição específica a ser utilizada para o ambiente da universidade, porque corre-se o risco do excesso tanto pela generalização, quanto pela particularização, permitindo facilmente confundir-se pela diversidade de temas que são abordados. Assim, estas definições aqui são adotadas como orientadoras, possibilitando sua referência conforme seja necessário e pertinente ao estudo. Contudo, sob a finalidade da observação da gestão dos dados no ambiente, deve ser destacado a amplitude relacional entre a RDM e a universidade.

Igualmente fundamentado na necessidade da amplitude destacada, este trabalho entende que a RDM deve ser tão abrangente ou tão restrita quanto o contexto abordado. Assim, é pertinente explorar a RDM nestes dois aspectos para o contexto da universidade, ou seja, abrangente o bastante para considerar as práticas aparentemente distantes da investigação, mas que com ela interage, ao mesmo tempo que deve observar as práticas do ambiente imediato do investigador. Neste sentido, o estabelecimento de uma RDM ampla e eficiente tende a ser mais complexa e desafiadora, pois quanto mais atenta-se para estes dois aspectos no ambiente, mais atores estarão envolvidos em um esforço para relacionar pontos aparentemente remotos.

No resumo da presente seção, importa concluir que esta visão relacional entre a RDM e a universidade indica que: 1) a universidade é desafiada a estar atenta aos dados de investigação administradas em seu âmbito; 2) importantes oportunidades, exigências e fatores evidenciam a necessidade da organização para a RDM na universidade; 3) as instituições enfrentam desafios, dentre os quais destacam-se a heterogeneidade, a coordenação dos elementos humanos e não humanos e os desafios técnicos e sociológicos; 4) embora haja desafios, mudanças estão em progresso na área para o avanço da RDM nas universidades; e, 5) a amplitude da RDM pode ser percebida em suas definições, direcionando para a abrangência e a especificidades, ambas merecedoras de igual atenção no ambiente. Assim, a visão relacional segue na próxima seção, agora direcionada pela identificação e classificação dos atores participantes em relação ao ambiente temático.

2.2. OS ATORES E AS ARTICULAÇÕES NO AMBIENTE

A relação entre a RDM e a universidade apresentada na seção anterior, indica o desafio para a organização dos atores humanos e não humanos. Assim, a presente seção, tem principalmente na pesquisa bibliográfica a colaboração para o reconhecimento de tais atores e suas articulações no ambiente, mas também a contribuição para a contextualização proposta neste capítulo. Portanto, o objetivo aqui é, através do que se tem produzido sobre o assunto, observar os atores envolvidos nas articulações para a RDM no ambiente da universidade, de modo a identificá-los para melhor gerir os dados da investigação.

Para identificar os atores e as articulações no ambiente temático, a pesquisa que segue realizou o levantamento da bibliografia e das informações contidas na bibliografia, bem como a síntese e análise dos textos (Lima & Miotto, 2007). Tal percurso foi efetivado em três etapas, a saber: 1) a coleta de dados; 2) a seleção dos textos; e, 3) a sintetização e análise dos conteúdos. Na primeira etapa – coleta de dados – adotaram-se critérios que delimitaram o universo da pesquisa, dentro dos parâmetros já detalhados na abordagem metodológica. A segunda etapa – seleção dos textos – foi definida através das informações presentes nas publicações recuperadas. A seleção se deu de modo exploratório, verificando a existência de indícios da relação entre os atores da RDM no título, resumo ou nas palavras-chave e, em uma leitura seletiva na introdução e conclusão para identificar uma possível articulação. A terceira etapa – sintetização e análise dos conteúdos – se deu nos textos das publicações selecionadas, de modo a identificar qual tipo de articulação a que se refere o documento. Tal procedimento originou à sintetização apresentada na tabela 7. Por fim, foi realizada a análise do conteúdo, comparando os resultados a fim de reconhecer os atores e os relacionamentos envolvidos nas articulações para a RDM no ambiente da universidade.

Com base nos procedimentos metodológicos utilizados, foram recuperados 70 artigos. Dentro dos critérios adotados e de acordo com as leituras específicas, 63 artigos foram eliminados, pois, ou estavam repetidos, ou não mencionavam as relações dos atores para a RDM na universidade de forma articulada. Assim, 7 das 70 publicações se apresentaram positivamente para os procedimentos supramencionados. Em consequência, as publicações selecionadas como definitivas e analisadas seguem na tabela abaixo.

Tabela 6*Publicações analisadas sobre a relação entre a RDM e a universidade*

Nº	Autor(es)	Título
1	Jamie Wittenberg, Mary Elings (2017)	Building a Research Data Management Service at the University of California, Berkeley.
2	Jochen Schirrwagen et al. (2019)	Expanding the Research Data Management service portfolio at Bielefeld university according to the three-pillar principle towards data FAIRness
3	Belinda Norman, Kate Valentine (2013)	From Project to Strategic Vision: Taking the Lead in Research Data Management Support at the University of Sydney Library
4	Ricky Erway (2013)	Starting the Conversation: University-Wide Research Data Management Policy
5	Ken Chadand Suzanne Enright (2014)	The research cycle and Research Data Management (RDM): Innovating approaches at the University of Westminster
6	Ulrich Meyer, Heike Neuroth (2015)	The Strengthening of Information Literacy in the Context of Research Data Management: a Challenge for Universities and Political Institutions
7	Wilson, Fraser, Martinez-Uribe et al. (2010)	Developing Infrastructure for Research Data Management at the University of Oxford

Ao examinar os títulos e os resumos das sete publicações selecionadas, pode-se perceber que quatro apresentam ênfases sobre serviços, uma publicação apresenta ênfase sobre política, uma sobre o ciclo da investigação e outra publicação apresenta ênfase sobre a literacia. Em uma primeira abordagem, foi possível verificar que os materiais recuperados tratam de forma destacada das políticas ou dos serviços. Outro fator realçado, é o ambiente principal da articulação da RDM, onde percebe-se que uns artigos se referem e se restringem à biblioteca e, outros, inserem a articulação da RDM no ciclo da investigação. De todo modo, é preciso observar mais detalhadamente os textos evidenciados na tabela a seguir.

Tabela 7*Referências que denotam a articulação entre atores da RDM e a universidade*

Nº	Referências extraídas no texto dos artigos que denotam articulação entre atores da RDM no contexto da universidade
1	<ul style="list-style-type: none"> ● University of California, Berkeley's Library and the central Research Information Technologies unit have collaborated to develop a Research Data Management program that leverages each organization's expertise and resources to create a unified service. ● (...) service areas that are often fully embedded in IT, like backup and secure storage, as well as services in the Library domain, like resource discovery and instruction, are integrated into a single Research Data Management program.

2	<ul style="list-style-type: none"> ● Research Data Management at Bielefeld University is considered as a cross-cutting task among central facilities and research groups at the faculties, Continuous Integration. ● The evolution of the institutional RDM is based on the three-pillar principle: 1. Policies, 2. Technical infrastructure and 3. Support structures. ● The problem of data quality and the issues with reproducibility of research data is addressed. ● It is creating an infrastructure for the processing and versioning of research data which will finally allow publishing of research data in the institutional repository.
3	<ul style="list-style-type: none"> ● The Library, ICT and Research Portfolio have their own distinct cultures, and much of the partnership development focused on articulating shared goals and establishing boundaries for departmental roles. ● The creation of strong partnerships across institutional units was one of the most valuable outcomes of the project, and will ensure that Research Data Management support will continue to be informed by collaboration between the University's research and teaching support units. ● (...)opportunity to articulate Research Data Management support roles and responsibilities within those faculties, in which the Library will take a leading role.
4	<ul style="list-style-type: none"> ● This call for action addresses the high-level benefits of adopting a university-wide policy regarding research data management. ● It is important to recognize the current uncertainty as to how data management support and services will be distributed among university, disciplinary, funder, national and international stakeholders. ● In this complex environment, an institution must actively determine how data services will be managed and distributed internally. ● Various university players are important stakeholders in determining the appropriate governance structure to ensure efficient coordination, adequate security and regulatory compliance, and scalable, sustainable, and useful data management services to researchers.
5	<ul style="list-style-type: none"> ● This article presents a case study based on experience of delivering a more joined-up approach to supporting institutional research activity and processes, Research Data Management (RDM) and Open Access (OA). ● The result of this small study, under taken at the University of Westminster in 2013, indicates that a more holistic approach should be adopted.
6	<ul style="list-style-type: none"> ● Coordinated action is also called for hereto ensure that existing bottom-up initiative scan develop into nationally and internationally harmonized Research Data Management. ● In order to implement research data management successfully at universities it is imperative that information literacy best lengthened appropriately. Students, teaching staff and researchers need to adapt to this new challenge, but so, too, do university service facilities. ● This is primarily the responsibility of the political institutions. It should be actively involved in the national coordination of Research Data Management.
7	<ul style="list-style-type: none"> ● The centrally-coordinated activities at Oxford have recognized these different priorities by trying to ensure that support for research data management is a distributed, 'multi-agency' activity ● The University of Oxford's institutional approach to research data management was built upon two principles: researchers being at the core of development; and the need for intra-institutional collaboration amongst service providers. ● Data management means different things to different people (...). It is this broadest definition of research data management that informs our approach to data management infrastructure.

A tabela 7 apresenta os excertos extraídos das publicações, que dão sentido ao conteúdo que foi verificado. Assim, a seguir tem-se um resumo da análise dos textos evidenciados, observando a articulação e os atores na relação entre a RDM e a universidade.

Texto n. 1. – A biblioteca e a equipe de Tecnologia da Informação (TI) são os atores principais desta articulação. O texto apresenta reflexões sobre a articulação entre atores complexos e díspares, com interesse na realização de uma atividade interdependente. Os serviços de armazenamento seguro e *backup* são citados em relação à TI, também a descoberta e a gestão de recursos com a biblioteca, todos integrados ao programa da RDM.

Texto n. 2. – Sobre o problema da qualidade e reprodutibilidade dos dados, o artigo propõe uma solução por intermédio da criação de uma infraestrutura para a expansão dos serviços para a RDM, que permitirá a publicação dos dados em repositório institucional e a integração contínua com os grupos de investigação direcionada para a reutilização dos dados. O artigo se restringe aos dados e aos serviços e elege três pilares de apoio, quais sejam: 1) as políticas; 2) a infraestrutura técnica; e, 3) as estruturas de apoio; porém, até por não ser o objetivo do material, não é desenvolvido o assunto no tocante aos pilares em questão.

Texto n. 3. – A articulação referenciada no texto situa-se entre a biblioteca, a TI e os investigadores, a partir da liderança da primeira, no sentido de apresentar o caso como uma evolução da biblioteca no suporte para a RDM. Com base nesta evolução, tal articulação é percebida como algo importante, mas não se observa nenhum detalhamento sobre o contexto articulatório, ou mais informações sobre esta relação. O texto ainda destaca a relevância das políticas e das práticas para a RDM.

Texto n. 4. – É uma chamada de atenção para a importância da adoção de políticas para a RDM em prol da universidade. Elege-se *stakeholders* e se propõem questões para o diálogo entre a universidade; *The Office of Research*; *The Research Compliance Office*; o departamento de TI; os investigadores; as unidades acadêmicas; e, a biblioteca. A publicação é mais abrangente em termos de amplitude para um possível diálogo articulatório entre atores diversos; porém, não se tem ali mais passos verificados além da proposição do diálogo e de sua importância para o estabelecimento de políticas para a RDM.

Texto n. 5. – Embora a palavra “holística” e a questão da abordagem conjunta para a gestão dos dados de investigação apareçam aqui mais de uma vez, o texto é um estudo de caso sobre uma ferramenta de informática e seus benefícios, desenvolvida para apoiar os investigadores durante o ciclo de vida da investigação.

Texto n. 6. – A articulação aqui evidenciada, apresenta-se na necessidade de ser estabelecida em níveis diversos, destacam-se as políticas para responder questões sobre as responsabilidades da biblioteca, dos investigadores e da TI. O texto aponta para tal necessidade, exemplificando casos de sucesso de políticas na Alemanha e reforçando a urgência de uma articulação coordenada dos principais atores da RDM no ambiente da universidade. Apesar de apontar para relevância de tal articulação em vários níveis, o artigo o faz apenas, com ênfase nas políticas.

Texto n. 7. – O documento tem sua centralidade na infraestrutura da RDM na Universidade de Oxford. O texto traz a necessidade de articulação, a partir da ideia de que a “gestão de dados significa coisas diferentes para pessoas diferentes” e procura realizar uma estratégia de infraestrutura coordenada e unificada, destacando a necessidade de adequação à RDM e a necessidade de colaboração. Apesar da ênfase na infraestrutura, a articulação em um ambiente macro da universidade é proposto. O artigo possui aplicação prática e auxilia no entendimento de como ocorre a articulação entre atores distintos.

Nos textos das publicações apresentadas, destaca-se a ênfase nas políticas e nas necessidades articulatórias entre os atores. No entanto, em sua maioria, reduz-se tal articulação a serviços em âmbitos específicos das universidades. Além da concentração nos serviços, têm-se apoios e suportes para a RDM no processo da investigação, as bibliotecas aparecem como órgãos centrais, ao passo que a TI também ganha destaque, se repetem ideias para a infraestrutura, ferramentas de *software*, DMP (*Data Management Plan*), princípios FAIR (*Findable, Accessible, Interoperable and Reusable*), curadoria e repositórios.

No intento de contribuir para a identificação dos atores para a RDM, verificou-se que em todo material levantado, eliminando-se os artigos repetidos, foram citados de forma recorrente, mas sem menção de articulação, os seguintes quantitativos sobre os respectivos assuntos: 14 textos tratavam de práticas e conhecimentos a respeito da RDM, para aplicação

na investigação da própria universidade; 15 textos tratavam sobre a implantação de serviços em geral, quase todos referentes à biblioteca da universidade; 2 textos tratavam de avaliações sobre serviços da RDM já implantados; e, 5 textos tratavam separadamente do seguinte: a relação entre a biblioteca universitária e a RDM, a construção de infraestrutura da RDM, uma proposta de DMP, fatores sobre a falta de incentivos por parte dos gestores da universidade e fatores sobre o engajamento institucional.

A finalidade de identificar atores, embora necessária, é uma questão de difícil realização pela própria característica da RDM no ambiente. Nessa direção Pinfield, Cox e Smith (2014) explicam que a RDM é uma questão complexa, que envolve várias atividades realizadas por muitos atores que abordam e são influenciados por um grande conjunto de fatores. Whyte e Tedds (2011) também alertam que permanecem questões em aberto sobre exatamente quem é o responsável e onde, no ecossistema de investigação.

O OpenAIRE (2017) colabora com o assunto ao afirmar que a RDM abrange múltiplas partes interessadas, quais sejam: o investigador e os assistentes de investigação; a administração da instituição; a biblioteca; a TI; os serviços jurídicos; os financiadores; os editores acadêmicos e curadores dos dados de investigação entre outros, e declara: “However different, these stakeholders all play a role in planning, carrying out, assessing and benefiting from good data management”; e, complementa que a RDM “[...] encourages heterogeneous stakeholder groups to work together for a shared societal goal” (p.10).

Jones, Pryor e Whyte (2013) também indicam três grandes grupos que devem ser observados para a RDM, a saber: 1) a gestão da universidade; 2) o suporte e serviços administrativos (biblioteca, TI, gerenciamento de registros e administração de investigação); e, 3) os investigadores. Os autores descrevem o assunto deste modo:

In the complex and frequently distributed organizational environment of an HEI (Higher Education Institutions), the roles and responsibilities for defining, enabling and delivering RDM services are generally shared across three groups: university management, support and administrative services, and researchers (p. 2).

Além de conhecer o contexto e os atores, é preciso que se tenha uma articulação entre eles e o ambiente da universidade. Whyte e Tedds (2011) apontam que em relação ao ambiente, é necessário identificar as áreas de responsabilidade da instituição e dos

investigadores e, que as universidades, necessitam comprometer-se com o desenvolvimento de diretrizes, treinamento, suporte adequados e implantação de serviços que exigem a cooperação e coordenação com outros atores. Jones, Pryor e Whyte (2013) também observam que as instituições de ensino superior precisam coordenar uma ampla gama de atores e processos para fornecer as infraestruturas tecnológicas e humanas necessárias, a fim de atender as necessidades da gestão dos dados.

Pode-se perceber pela pesquisa bibliográfica, mas também nas referências até aqui apontadas, que o ambiente da RDM na universidade traz uma série de termos, assuntos, práticas e atores que precisam ser organizados para um melhor entendimento das conexões, articulações e ações decorrentes. Portanto, pode-se inferir desta seção, a proposição de três tipos de atores que podem servir como base para agrupar os demais, são eles: 1) as pessoas; 2) as políticas; e, 3) as infraestruturas. Esta proposta é fundamentada nos seguintes fatores: a) a necessidade da categorização para melhor organização; b) a importância destes meta-atores inicialmente referenciados; e, c) a recorrência em que estes atores são aqui citados.

Quanto à organização, a tríade sugerida parece poder estruturar e organizar em si, grande parte dos termos, dos assuntos, das práticas e da diversidade de atores que dizem respeito à RDM. Esta organização é importante porque está no cerne do aprendizado. Nesse sentido, Harnad (2005) diz que “cognição é categorização”, ao passo que Taylor e Joudrey (2008) acrescentam que o ato de organizar serve para compreender as conexões existentes. Quanto à importância, os três atores sugeridos se apresentam de forma significativa, pois, a partir das análises dos resultados da pesquisa bibliográfica e das referências, percebem-se que as citações a tais atores são realizadas de forma assídua (não necessariamente com estes mesmos nomes).

Ao caracterizar os atores aqui agrupados, pode-se perceber que o primeiro ator – *pessoas* – refere-se ao relacionamento (indivíduos/instituições) que é parte da abordagem sociológica no ambiente, tendo em vista a possibilidade de surgirem barreiras comportamentais e contextuais que impedem uma maior adesão dos elementos humanos à RDM (Jones, Pryor & Whyte, 2013). No contexto proposto, percebe-se um intercâmbio social, onde as pessoas interagem entre si e com recursos, estruturas e processos, o que significa que essas interações não somente se apresentam como barreiras, mas trazem também benefícios ao contexto (Veiga, 2017). Portanto, aspectos positivos e negativos podem ser

analisados através dos relacionamentos de interesse e na relação de influência das pessoas com os outros atores. Assim, gestores, investigadores, comunidade acadêmica, profissionais da informação e outros agentes aparecem no ecossistema da RDM, de forma proeminente nas articulações, para o avanço ou como impedimento para a gestão dos dados na universidade.

O segundo ator – *políticas* – tem seu destaque por ser um dispositivo que influi diretamente na coordenação e nas práticas para a RDM em todos os níveis da instituição. Grande parte do material recuperado na pesquisa bibliográfica, e aqui referenciado, trata sobre as políticas que envolve a RDM em variados níveis de abrangência e amplitude. Sobre a questão, Shearer (2015) destaca que as políticas são uma alavanca extremamente poderosa, pois fornecem uma estrutura que ajuda a orientar as melhores práticas e que, sem elas, seria improvável uma ampla adoção da RDM. Asher et al. (2013), ratificam que a falta de políticas institucionais é um dos fatores que dificultam o progresso da RDM. Portanto, as políticas têm um papel essencial no ambiente, pois são instrumentos de integração entre atores, clarificando papéis e responsabilidades em como atuar para a RDM na universidade.

As *infraestruturas*, como terceiro ator, normalmente são percebidas como algo físico, que se pode ver. Nesse sentido, Slota e Bowker (2017), atentam que, em geral, a percepção da infraestrutura:

[...] refers to the prior work (be it building, organization, agreement on standards, and so forth) that supports and enables the activity we are really engaged in doing. More particularly, infrastructure refers to those systems, technologies, organizations, and built artifacts (p. 1).

A partir desse entendimento tradicional, a infraestrutura é referida em muitos objetos de estudo no ambiente da RDM, tais como: repositórios institucionais, tecnologias, comunicação, treinamento, construção de DMPs, além de variados serviços e princípios associados. Assim, as infraestruturas podem incluir não apenas toda a parte técnica e tecnológica, mas outros artefatos a serem desenvolvidos para a RDM.

A articulação entre os três atores sugeridos, para este estudo, não significa ligação entre eles, mas um equilíbrio ou coordenação quando direcionados à RDM na universidade. Shearer (2015), por exemplo, aponta esse equilíbrio como essencial, quando justifica através do modelo conhecido como “*three-legged tool*” (Kenney & McGovern, 2005), a relevância das políticas para a RDM, no equilíbrio entre a cultura, a infraestrutura e os serviços. No

mesmo sentido, Brown, Bruce e Kernohan (2015), desenvolvendo uma visão de futuro na RDM para as universidades do Reino Unido, observam o seguinte sobre atores em equilíbrio: “[...] it simply takes time to put infrastructure – policy, people and technology – in place” (p. 6). Achard et al. (2013), também destacam a necessidade de uma abordagem coordenada entre atores:

It is also important to note the linkage between Research Data Policy, Technology and Support. To promote conscious and successful use of research data, these three aspects should be offered simultaneously to researchers. Projects that merely focus on one or two of these aspects are doomed to fail, as well as projects where policy, support and services are not aligned. A coordinated and parallel approach is therefore crucial (p. 3).

A necessidade da articulação supramencionada é ratificada na pesquisa bibliográfica, quando se verifica que as práticas para a RDM, as políticas institucionais, repositórios, bibliotecas, curadoria, DMP, princípios FAIR e outros, são tratados de forma individual, realçando assim um desequilíbrio, ao mesmo tempo que é destacada a necessidade do relacionamento entre tais atores. Nesse sentido, Brown, Bruce e Kernohan (2015), concluem que são fundamentais ações coordenadas para uma RDM mais eficiente e eficaz.

Diante do exposto, é possível pensar que os três atores aqui sugeridos – pessoas, políticas e infraestruturas - quando articulados podem permitir agrupar e coordenar de modo mais organizado vários elementos relacionados à RDM na universidade. No entanto, aglutinar tais elementos entre os três atores sugeridos pode ser aparentemente tarefa mais simples para uma compreensão teórica inicial. Porém, quando o objetivo maior é desenvolver um modelo para articulação em um ambiente específico, com características próprias como no caso das Unidades de Investigação e Desenvolvimento, é possível que os meta-atores sugeridos e agrupados nesta seção não abarquem, ou não sejam adequados, para atender de forma ampla as especificidades para a RDM no ambiente da universidade. Portanto, os meta-atores aqui identificados devem ser entendidos de acordo com um dos propósitos deste capítulo, ou seja, como uma fundamentação e contextualização inicial para uso e direcionamento do estudo.

O resultado desta seção aponta que: 1) há uma gama complexa e díspar de atores no ambiente da RDM da universidade; 2) é necessário organizar tais atores no ambiente; 3) é possível agrupar diversos atores da RDM na tríade pessoas-políticas-infraestruturas, mas esta tríade deve ser vista como uma orientação inicial; e, 4) é preciso que os atores se equilibrem em articulações necessárias e ações coordenadas no ambiente da universidade.

2.3. AS CARACTERÍSTICAS GERAIS DOS DADOS DE INVESTIGAÇÃO

A relação entre os dados de investigação e a universidade pode ser vista sob muitas perspectivas. Para os objetivos desta investigação é preciso conhecer os dados como o alvo da gestão, de modo que este objeto informacional seja também percebido na perspectiva do seu valor, importância e principalmente através das suas características, para possibilitar uma relação pertinentes com a RDM na universidade. Entende-se que assim, os atores envolvidos na RDM podem melhor gerir os dados no ambiente.

Em uma perspectiva de importância, já em 2010, Borgman anunciava um “dilúvio de dados”, dois anos depois a autora declara que esse dilúvio previsto pela comunidade científica havia chegado (Borgman, 2010b, 2012). Tal fenômeno é atribuído aos avanços das tecnologias em mais de um século de progressos na área (Kelleher & Tierney, 2018). Assim, é possível pensar que as tecnologias não pararam de evoluir e, em consequência, o volume dos dados também não tende à diminuição. Nesse sentido, questões que permeiam os dados e a RDM se tornam mais atuais e sem perspectiva de arrefecer.

Quanto ao valor dos dados, este valor não é destacado apenas na atualidade, pois governos, empresas e a ciência sempre produziram dados, permitindo chegar a conclusões e decisões importantes em diversos segmentos. No entanto, o que hoje muda é o volume que se tem à disposição (Borgman, Wallis & Enyedy, 2007) e as diversas possibilidades de aplicações que podem ser realizadas para ampliação do seu valor. O problema é que mesmo sabendo que são valiosos, seja para a ciência ou para a sociedade, dado o volume e a rapidez com que são gerados, muitas vezes não se sabe o que fazer com eles e nem como gerir de forma a potencializar seu uso.

Na área da ciência, a grande produção de dados e a possibilidade de maximizar sua utilidade, juntamente com a perspectiva de desenvolvimento das ferramentas e políticas para acompanhar as transformações no ambiente científico, fizeram com que termos como *data-driven*; *e-science*; *data-intensive computing*; *cyberinfrastructure*; *fourth paradigm of science*; *data deluge*; e, *E-infrastructure* (Appelbe & Bannon, 2007; Frické, 2015; Gray, 2007; Hey & Trefethen, 2003) fossem descritos por diversos autores para indicar o surgimento de uma ciência intensiva em dados. Nesse contexto, o “quarto paradigma”, por

exemplo, remete a novas e diferentes formas de fazer investigação e novos modos de comunicação acadêmica, baseado no uso de grandes conjuntos de dados e tecnologias de computação (Hey, Tansley & Tolle 2009). Autores como Gray et al. (2005), e Taylor (2001), indicavam que o resultado da exploração, análise e integração de grandes volumes de dados possibilitam novas análises em colaboração global, tendo por base uma infraestrutura propícia para efetivá-la. Sob tal perspectiva, os dados nas universidades encontram seu lugar para que expectativas semelhantes a estas possam ser viabilizadas.

Em relação à universidade, os dados de investigação podem ganhar escopos científicos próprios, com características bem definidas. Assim, Borgman et al. (2009, 2016), Heidorn (2008), Wallis et al. (2013), explicam que os dados produzidos e utilizados pela investigação são caracterizados por duas dimensões, quais sejam: 1) *Big Science* – um pequeno segmento da ciência que envolve grandes equipes e orçamentos, com projetos duradouros, altamente instrumentados e complexos, mas consistentes em sua estrutura e produtora de grandes volumes de dados distribuídos em escala mundial; e, 2) *Small Science* – conhecida como a “cauda longa” (*long tail*) da ciência, que é constituída por indivíduos e pequenas equipes, em projetos específicos, onde uma grande quantidade de pequenas coleções de dados são coletados e gerados de muitas e diferentes formas, por uma infinidade de laboratórios de universidades e institutos de investigação, com alto valor informacional em todas as áreas do conhecimento.

No cenário em questão, a universidade pode ser vista sob as duas dimensões descritas, pois a depender da sua capacidade científica, é capaz de produzir dados com características semelhantes à da *Big Science*. No entanto, é mais comum que os dados produzidos na universidade estejam no escopo da “cauda longa” da ciência. Nessa perspectiva, segundo o e-IRG (2016) e Hedstrom e Myers (2017), os dados em uso na “cauda longa” estão presentes em todas as disciplinas, mas fragmentados e mantidos em milhões de arquivos de computadores individuais e servidores das universidades, acompanhados de um conjunto mínimo ou inexistente de documentação. São dados que, embora valiosos, são difíceis de descobrir e utilizar, por conseguinte, têm menor probabilidade de serem estruturados e de permitir a transferência e a disponibilidade entre equipes ou indivíduos, recaindo quase sempre no desuso e no esquecimento.

A importância de tal contexto é destacado por Heidorn (2008), quando afirmou: “It seems likely that transformative science is more likely to come from the tail than the head” (p.282). Nesse sentido, ainda existe um campo independente e fértil para o desenvolvimento de novas ideias e para a concepção de uma ciência insólita. Wyborn e Lehnert (2016) ratificam esse entendimento quando expressam que a organização dos dados no escopo em questão:

[...] have the potential to generated new knowledge. In other words, small data, when properly curated, can be compared and integrated to reveal large-scale temporal and spatial patterns that could lead to new scientific discoveries and insights (p. 1).

No entanto, embora a “Small Science will generate two to three times more data than Big Science [...] is horribly heterogeneous and far more vast” (Carlson, 2006, para. 12). Também Ferguson et al. (2014), alertam que a diversidade que caracteriza as grandes quantidades de pequenas coleções de dados é, ao mesmo tempo, rica e importante em informação e obstáculo para a sua disponibilidade e gestão. Heidorn (2008) também observa que em tal ambiente, frequentemente, os dados “[...] exists only in the bottom left-hand desk drawer of scientists on some media that is quickly aging and soon will be unreadable by commonly available devices” (p. 281). Tem-se, portanto, um problema de invisibilidade no que tange às coleções dos dados geradas no segmento da “cauda longa” da ciência e, neste sentido, no ambiente de investigação das universidades (Borgman, 2018). O e-IRG (2016) e Hedstrom e Myers (2014) resumem que os dados em questão são complicados para gerenciar e partilhar, pois estão por todos os lugares, mas, se explorados apropriadamente, podem otimizar o potencial da ciência. Assim, os dados de investigação têm emergido como um grande desafio e as possíveis soluções se tornam mais urgentes.

Na busca de viabilizar possíveis soluções, a tentativa de definir os “dados de investigação” é uma necessidade a ser realizada. No entanto, o ambiente complexo e heterogêneo da universidade e a diversidade dos próprios dados, fazem com que tentar padronizá-los sob uma definição seja uma tarefa complicada. Segundo Borgman (2010), uma definição consensual apresenta dificuldades, pois a identificação dos dados é passível de interpretações e reinterpretções, isto porque seus significados e amplitudes se transformam com as áreas científicas e domínios específicos a que estejam subordinados.

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE, 2007) descreve os dados como “(...) factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings” (p. 8) – descrição que pode ser combinada com a natureza interpretável e reinterpretável dos dados e dependentes de documentação e contexto (Consultative Committee for Space Data Systems [CCSDS], 2002). Neste sentido, Gray et al. (2002), descrevem os dados como incompreensíveis e, portanto, inúteis, a menos que se tenha uma descrição detalhada e clara de como e quando foram obtidos e de como foram produzidos.

Observado sob uma perspectiva generalista e sem expressar a diversidade contida nos dados, a combinação acima se adequa a uma definição possível de ser adotada, no sentido de que os dados de investigação podem ser vistos como objetos informacionais aceitos pela comunidade científica para a validação dos resultados da investigação, mas que devem também conter os meios necessários para serem caracterizados e (re)interpretados. No entanto, se uma identificação consensual apresenta dificuldades e uma definição geral não consegue conter os matizes dos dados no ambiente, é importante buscar o que pode ser atribuído como “propriedades” para que se possa caracterizar os “dados de investigação”.

No trabalho desenvolvido por Sayão e Sales (2020), com o objetivo de estabelecer uma base terminológica funcional para os dados de investigação, os autores sugerem a adoção de uma taxonomia que permite servir de parâmetro para a caracterização dos dados (Tabela 8). Tal caracterização pode ajudar a identificar e a gerir os dados no ambiente multidisciplinar da universidade, levando em consideração as especificidades trabalhadas na investigação e as infraestruturas tecnológicas e organizacionais direcionadas à sua gestão. Outros aspectos nos quais a caracterização dos dados pode ser necessária, estão na verificação do valor e/ou importância dos dados, mas também o desenvolvimento de políticas e preparação dos recursos humanos para tratamento no âmbito da RDM. A conjunção destes e outros fatores é capaz de viabilizar a RDM de acordo com a complexidade e as características dos dados em áreas distintas. Segue na tabela abaixo a taxonomia referenciada no trabalho de Sayão e Sales (2020).

Tabela 8*Proposta de taxonomia dos dados de investigação*

Taxonomia da investigação		
Em relação	Tipo	Subtipo
À origem	Dados de investigação	Observacionais, experimentais e computacionais
	Dados para a investigação	Governamentais, registros e dados transacionais
Ao grau de processamento	Dados brutos/primários/crús	
	Dados intermediários/pré-processados/secundários	Dados limpos, dados analisados, dados anonimizados e dados derivados
	Dados finais/processado/terciário	Dados publicados/arquivados e dados de alta densidade
Ao tipo de coleta	Automática	
	Manual	
Ao nível de sensibilidade	Dados não sensíveis/ostensivos	
	Dados confidenciais	
	Dados pessoais	Dados pessoais anonimizados e dados pessoais sensíveis
À abordagem	Dados qualitativos	
	Dados quantitativos	
À materialidade	Dados digitais	
	Dados físicos	Amostra, espécime e artefato
À perenidade	Dados episódicos	Institucionais, comunitárias e individuais
	Dados canônicos/referenciais	
À abertura	Grau de abertura	Fechado, aberto, parcialmente aberto e embargado
	Tipo de licença para uso	Creative Commons e outras possíveis
À natureza	Número	Medidas, resultado de levantamento, fórmula, equação e algoritmo
	Multimídia	Imagem, vídeo, áudio, animação, filme e fotografia
	Software	Base de dados, simulação e código
	Visualização	Tabelas, gráficos, diagramas, modelo em 3D, modelo reduzido e desenho
	Textual	Metadados, questionário, entrevista, anotação, certificado, caderno de laboratório, transição, correspondência, diário e caderno de campo
	Artefato	Espécime, amostra, maquete, phantom/manequim
	Processo	Procedimentos operacionais padronizados, workflow, protocolo e teste

Nota. Adaptado de Sayão e Sales (2020).

A partir da tabela anterior percebe-se que os dados de investigação podem ser caracterizados sob diversos aspectos. Assim, é necessário verificar os atributos sob a aplicação da RDM na universidade. No entanto, de modo semelhante à seção anterior onde os atores necessitam ser identificados de acordo com as especificidades, aqui os dados a serem observados também dependem da observação dos ambientes disciplinares específicos

da universidade. Portanto, deve-se conhecer o contexto para a caracterização dos dados, de forma que possibilite diminuir os obstáculos para a sua gestão, facilitando a estruturação e maximizando os benefícios que estes ativos informacionais podem proporcionar.

Com base nesta seção, faz-se considerável salientar que: 1) o aumento do volume dos dados impacta a investigação científica e que estes são importantes para a ciência e para a sociedade; 2) a maior produção e o potencial de uso dos dados está na chamada “cauda longa” da ciência, desenvolvida na universidade; 3) os desafios para identificar, organizar e gerir os dados nas universidades apontam para a heterogeneidade, com dados distribuídos em todas as disciplinas, difícil de estruturar e descobrir, com pouca documentação, tendendo a perdas e a invisibilidade; e, 4) os dados possuem características diversas, que precisam ser observadas de forma específicas e contextual para possibilitar uma melhor RDM na universidade.

2.4. CONSIDERAÇÕES SOBRE A GESTÃO DOS DADOS NA UNIVERSIDADE

O presente capítulo contextualizou teoricamente o ambiente temático da RDM na universidade, através de uma visão relacional direcionada pelo primeiro objetivo deste trabalho. Assim, foi realizado um panorama sobre a relação entre a RDM e a universidade e sugerida a identificação dos atores através de agrupamentos. Foram também apresentadas as características gerais dos dados de investigação. Obtiveram-se ainda, os parâmetros para ajudar na resposta da questão complementar: quem são os atores da RDM na universidade? Portanto, com base na revisão da literatura e na pesquisa bibliográfica, segue a síntese do capítulo.

Foi aqui indicado que as universidades, como produtoras de grandes quantidades de dados, têm em seu âmbito fatores que exigem a organização e práticas pertinentes para a RDM. Estas instituições, bem como os investigadores, enfrentam desafios técnicos e sociológicos no ambiente heterogêneo da RDM na universidade. Nesse sentido, o entendimento sobre a RDM aponta para uma amplitude de características, com práticas abrangentes e generalistas, que podem se relacionar com práticas específicas e localizadas. Nessa dualidade, ambas são merecedoras de igual atenção no ambiente da universidade.

O referencial do estudo indica que o ambiente da universidade é de difícil análise, pois se tem ali uma gama complexa e heterogênea de atores no contexto da RDM. Sob tais características ambientais, são estimuladas ações articuladas na assertiva de que é preciso

identificar, organizar e coordenar os atores do ambiente. Nesse sentido, foi reconhecido como classes de atores importantes e necessários: as pessoas, as políticas e as infraestruturas, e sugerido que tais classes possibilitam um modo de organizar em si os atores da RDM no ambiente. Apesar disso, é necessário perceber que estes atores não são capazes de abranger todos os participantes, nem a gama de especificidades constantes no ambiente da RDM na universidade, demandando assim uma melhor e posterior análise para identificar os atores observados a partir do ambiente da investigação na universidade.

Sobre os dados, é importante afirmar que estes crescem em volume, variedade e são relevantes tanto para a sociedade quanto para a ciência. Estes aspectos vem transformando a investigação científica, com novas formas de análise, colaborações e no modo do fazer científico, sendo que a maior produção e uso potencial dos dados está no segmento chamado de “cauda longa” da ciência. Os dados produzidos nesse segmento, na qual as universidades podem ser incluídas, são carregados de valor e possuem tipologias diversas, mas também são identificados como heterogêneos, fragmentados e “invisíveis”. Portanto, as características diversas dos dados precisam ser observadas, de forma específica e contextual para sua melhor gestão no ambiente. Infere-se que a superação dos desafios para a identificar, gerir e utilizar os dados na universidade, podem minimizar tais problemas.

Para resposta da questão complementar, um caminho possível apontado para identificar os atores da RDM na universidade são as classes genéricas, como sugerido na tríade de atores – pessoa, políticas e infraestruturas. Contudo, por se tratar de um ambiente complexo e com dados heterogêneos, a definição de tais atores depende de um contexto específico a ser explorado, considerando que os elementos humanos e não humanos podem muito variar entre áreas disciplinares diferentes. Neste sentido, é preciso observar o contexto para identificar quais são os atores para a RDM no ambiente.

A contextualização da RDM na universidade e os elementos teóricos apresentados neste capítulo, têm seu uso no decorrer deste trabalho, com desdobramentos que melhor fundamentam a resposta da questão levantada e forma organizada. Assim, a partir desta visão geral, os capítulos seguintes responderão as outras questões, objetivado pela estruturação de um modelo para a RDM que possibilite uma partilha adequada dos dados no ambiente da universidade, viabilizado pela articulação de atores e por práticas essenciais para a RDM.

3. CIÊNCIA ABERTA - UMA PARTILHA ADEQUADA DOS DADOS DE INVESTIGAÇÃO

A Ciência Aberta (CA) é basilar para este estudo da RDM (*Research Data Management*) na universidade, isto porque a CA possui um ferramental teórico composto de princípios e práticas, bem como elementos capazes de ajudar a responder à questão levantada e alcançar o objetivo deste trabalho, especialmente no que envolve a partilha dos dados de investigação. Assim, a CA, direcionada pelos seus princípios e práticas, é desenvolvida no presente capítulo em busca da resposta para a questão complementar: o que é uma partilha adequada dos dados de investigação? Para ser mais específico na pergunta e melhor guiar a construção da resposta, a questão proposta é reformulada da seguinte forma: o que deve ser considerado para realizar uma partilha adequada dos dados de investigação, no ambiente da universidade, sob o contexto da ciência aberta? Portanto fundamentado na CA, o capítulo busca reunir os elementos necessários para responder à questão e tornar possível a caracterização de uma partilha adequada dos dados.

No direcionamento acima apresentado, este capítulo é composto por seis seções, quais sejam: 1) a partilha em acesso aberto; 2) os princípios e práticas da ciência aberta; 3) uma proposta de organização relacional para os princípios e práticas da CA; 4) as tensões e conflitos para a partilha dos dados; 5) a abertura dos dados de investigação; e, 6) as considerações finais para realizar uma partilha adequada dos dados de investigação.

As cinco seções iniciais foram constituídas para primeiro perceber a importância e as características da partilha em acesso aberto, depois organizar de modo relacional os princípios e as práticas da CA que podem orientar a partilha dos dados em acesso aberto, em seguida e sob o mesmo contexto, considera-se o comportamento dos investigadores na forma de tensões e conflitos para a partilha e, então, destaca-se uma visão geral para a abertura dos dados de investigação. O capítulo termina com as considerações finais, que possibilitam através dos produtos das seções, responder à questão e caracterizar uma partilha adequada dos dados de investigação. Permite-se assim, cumprir o segundo objetivo específico deste trabalho.

3.1. A PARTILHA EM ACESSO ABERTO

A ciência aberta tem como um dos seus pilares o Acesso Aberto (AA). Em termos gerais, o AA pode ser entendido como a disponibilização da produção científica, de forma pública, digital, *online*, irrestrita e gratuita ou, com um menor número de restrições e menor custo possível (Budapest Open Access Initiative [BOAI], 2002; Suber, 2012). No mesmo sentido, o termo “aberto” é passível de ser percebido em relação a abertura⁹ dos dados, do conteúdo e do conhecimento, de modo a ser assim considerado “if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness.” (Open Knowledge Foundation, s.d., para.2). Assim, através do processo de abertura, a partilha em AA possibilita a disponibilização da produção científica de modo aberto. Em relação a partilha dos dados em AA¹⁰, a abertura é desenvolvida no processo da investigação e tem como resultado os dados abertos.

O contexto da CA pode induzir que os “dados abertos” e o AA sejam interpretados sob o mesmo significado, ou que o primeiro seja uma vertente do segundo. No entanto, deve ficar claro que “dados abertos” se refere a partilha pública dos dados de investigação e envolve um espectro amplo de categorias e aplicações no campo da CA (Albagli, Clinio & Raychtock, 2014), e onde tais dados ganham destaque para a partilha em AA por serem objetos informacionais distintos dos demais elementos da produção científica, em termos de complexidade, diversidade e dos seus significados.

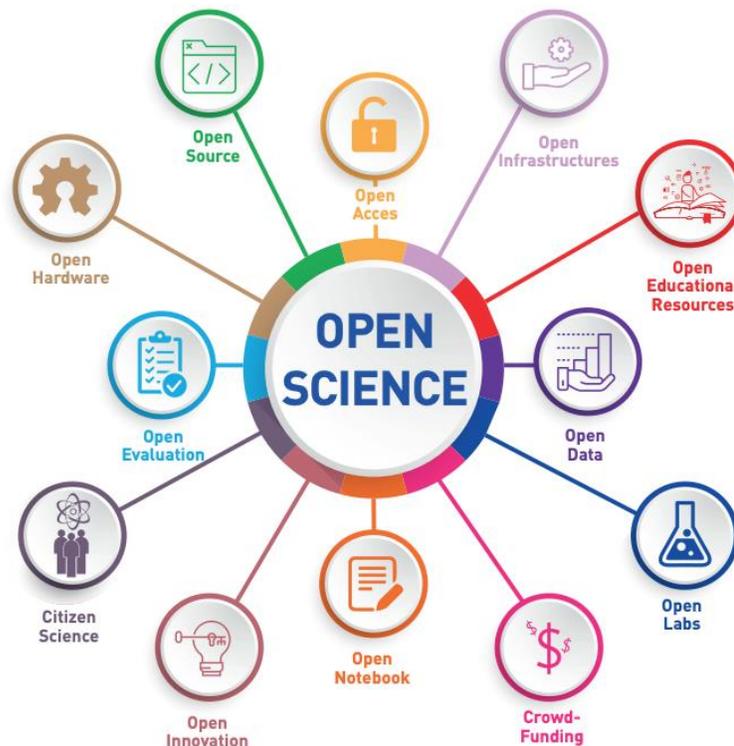
A figura abaixo ajuda a perceber que a CA envolve também outros componentes que requerem ser abertos, onde neste contexto, o AA e os dados abertos embora estejam teoricamente relacionados, são componentes distintos da CA e portanto, apresentam problemas e características específicas que devem ser assim observadas. Assim, embora diferentes, a relação entre a CA, a abertura (dos dados) e o AA estão conceituadas de forma interdependente (Willinsky, 2005), de modo que o construto para a partilha em AA pode ser aplicado à partilha dos dados em AA, obviamente respeitando as particularidades encontradas.

⁹ A “abertura” refere-se ao processo para que o objeto informacional e o acesso ao conhecimento se torne “aberto”.

¹⁰ Neste trabalho, há distinção entre o termo “partilha em acesso aberto” e o termo “partilha dos dados em acesso aberto”, o primeiro refere-se à disponibilização da produção científica, incluindo os dados de investigação, já o segundo remete exclusivamente aos dados e a sua “abertura” na investigação.

Figura 3

O acesso aberto e os dados abertos como componentes da ciência aberta



Nota. United Nations Educational, Scientific and Cultural Organization (Unesco, 2021).

Para que a disponibilização dos dados de investigação e dos demais elementos da produção científica, bem como o acesso ao conhecimento estejam abertos de acordo com os termos referenciados, é necessário o ato da partilha. Parece óbvio afirmar que há necessidade do ato da partilha para a disponibilização em AA, mas é justamente por isso que a ênfase deve recair primeiramente sobre esta ação, pois o modo como a partilha é realizada tem influência e consequência direta no cumprimento dos objetivos do AA. Por sua vez, os objetivos do AA viabilizados pela partilha são diversos e vão desde a redução da desigualdade informacional (Bezuidenhout et al., 2017; UNESCO, 2021b), até a identificação de soluções para os desafios globais da contemporaneidade, como no caso da pandemia da COVID-19. Nesse sentido, a UNESCO¹¹ (2021b) através da orientação: *Open access to facilitate research and information on COVID-19*, com objetivo de prevenir e

¹¹ Em 30 de março de 2020, a UNESCO organizou uma reunião online com representantes dos Ministérios da Ciência de 122 países, para trocar opiniões sobre o papel da cooperação internacional em ciência e aumentar o investimento no contexto da COVID-19.

mitigar crises a níveis globais, pediu aos governos que reforçassem a cooperação científica e integrassem a CA em seus programas de investigação. A solicitação destaca a necessidade de melhorar o acesso aos dados e à informação científica em países que não os possuam, tratando a partilha em AA como essencial na elaboração de respostas científicas, motivados pela urgência de soluções, para lidar com a crise pandêmica ocasionada pela COVID-19¹²:

In the face of the Covid-19 pandemic, the value and necessity of Open Solutions is crucial. Open Access to scientific information and open data facilitate better and faster research towards a vaccine and inform public health measures essential to contain the spread of the virus (para.1).

A importância e a influência da partilha em AA podem assim ganhar dimensões globais. Porém, é preciso entender que tais dimensões fazem parte de um sistema progressivo e histórico de partilha do conhecimento a partir da investigação local. Ou seja, a perspectiva da realização de soluções globais estão intrinsecamente ligadas a práticas específicas que envolvem os investigadores e o processo de investigação.

A definição seminal da *Budapest Open Access Initiative* (BOAI) e a explicação da Comissão Europeia sobre o AA, ajudam a perceber que a prática da partilha é inerente às ações no âmbito da investigação e do investigador, pois fazem referência aos direitos, restrições e as informações científicas a serem disponibilizadas em AA, sob práticas no ambiente académico e da comunicação da ciência. A Comissão Europeia (2016) assim contribui para o entendimento do contexto dos elementos e da partilha em AA:

Open access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean: 1. peer-reviewed scientific research articles (published in scholarly journals), or 2. research data (data underlying publications, curated data and/or raw data) (p.2).

¹² Várias foram as iniciativas e apelos para a realização de soluções abertas para a informação científica, por ocasião da crise pandêmica da COVID-19, como por exemplo, a declaração: *A question of survival - and example for an open Society*, do Comitê Executivo da Comissão Alemã da UNESCO, que enfatiza a ciência aberta como uma questão necessária para a superação da pandemia. Os assessores governamentais de ciência de doze países divulgaram uma carta aberta (*open letter*), pedindo a editores científicos que tornassem todas as investigações relacionadas à COVID-19 disponíveis de forma mais livre, por meio de fontes como o banco de dados COVID da Organização Mundial da Saúde. A Coalizão Internacional de Consórcios de Biblioteca (ICOLC) emitiu uma declaração (*Statement on the Global COVID-19 Pandemic and Impact on Library Services and Resources*), para ajudar as editoras e provedores de conteúdo a compreender o quanto a COVID-19 afeta a comunidade da informação. As Data Together Organisations (CODATA, GO FAIR, RDA, e WDS), em reconhecimento à necessidade de respostas para o acesso aos dados e políticas durante tempos de crise, lançaram o apelo *COVID-19 Appeal and Actions*.

A BOAI (2002), explica a finalidade do AA nos seguintes termos:

[...] free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited (para 3).

As citações anteriores, além de ratificarem a ideia de que a abertura para a partilha começa na investigação local, também podem representar as características fundamentais do AA que permeiam a literatura da área, em termos de filosofia aberta e em como o material, incluindo os dados devem ser partilhados, ou seja, de forma: a) *online* – por conexão direta ou remota a um computador ou a uma rede de computadores, como a Internet; b) gratuita e sem – ou com mínimas - barreiras, como as financeiras, jurídicas ou técnicas; c) livre para uso, por qualquer pessoa e para qualquer finalidade, respeitados os direitos, como os de reconhecimento e citação do autor; d) a ser representada por artigos e dados de investigação; e) que as informações científicas disponíveis sejam reutilizáveis; e, f) no contexto da investigação e inovação, para todas as disciplinas acadêmicas.

Portanto, para que os dados sejam partilhados sob tais características referenciadas para o AA, os investigadores, produtores e detentores dos objetos informacionais são solicitados a partilhá-los como parte do processo de investigação sob diversas justificativas, entre elas o de tornar possível o uso dos dados para além dos fins aos quais foram inicialmente produzidos e que estejam “[...] accessible and usable to anyone, anytime, anywhere, and for any purpose” (Faniel & Zimmerman, 2017, p.59). E ainda, quando as investigações são financiadas com recursos públicos, muitas agências financiadoras requerem que os resultados da investigação sejam disponibilizados publicamente em AA. São solicitações e às vezes, imposições que recaem sobre os investigadores, direcionados em abordagens do tipo *bottom up* e *top down* para partilharem em AA (Borgman, 2012).

É no sentido acima exposto que o ato ou o modo como a partilha é realizada tem influência e consequência direta no cumprimento dos objetivos do AA, pois embora a partilha em AA seja necessária e as abordagens, justificativas e as exigências decorrentes estejam no âmbito do trabalho do investigador, é importante atentar para que o ato da

partilha não se torne um fim em si mesmo (Pasquetto, Randles & Borgman, 2017) e percebida pelo investigador apenas como parte da rotina no processo de investigação para o cumprimento de obrigações. Isto porque a ação da partilha é um procedimento relativamente simples envolto em uma atividade com elevado grau de complexidade como a investigação científica. A simplicidade encontra-se no próprio ato, considerando que a ação da partilha pode ser entendida como “[...] the release of research data for use by others.” (Borgman, 2012, p.4). Assim, e por motivos diferentes, pode-se realizá-las de várias formas, incluindo os métodos informais ou *ad hoc*, como por exemplo, a troca privada de dados entre investigadores ou a disponibilização em *sites* de Internet, *pen drives*, e o depósito em computadores ligados em redes (Fienberg, Martin & Straf, 1985; Kim, 2013).

Por outro lado, em um contexto atento aos objetivos do AA, a partilha envolve princípios e práticas que influenciam diretamente na qualidade do processo da própria partilha e dos dados partilhados. O problema é que os processos e as práticas para partilhar em AA são muitas vezes extensas, suscitam dúvidas, questionamentos e requerem habilidades que transpõem a experiência de muitos investigadores. Some-se que os desincentivos geralmente superam os incentivos (Borgman, 2015; Mayernik, 2016; Pasquetto, Randles & Borgman, 2017; Wallis et al., 2013). Assim, as dificuldades para partilhar em AA podem influenciar e criar impedimentos para que esta seja realizada de modo adequado, tais dificuldades são possíveis de ocorrer por questões individuais e contextuais, barreiras técnicas, sociais, culturais e psicológicas (Tenopir et al., 2015; Zhu, 2020).

Nesta seção, de modo introdutório, põe-se em destaque as características, a importância e também algumas dificuldades para a partilha em AA, tais fatores quando pensados para o ambiente dos dados de investigação na universidade, permite produzir o seguinte raciocínio: se a partilha dos dados em AA, embora importante e necessária, pode acrescentar dificuldades ao já complexo processo da investigação e, se as imposições e solicitações para partilhar os dados em AA podem tornar o ato da partilha um fim em si mesmo, então, é necessário verificar o que deve ser realizado para reduzir tais dificuldades. Neste sentido, as próximas seções além de ajudarem a responder o que deve ser considerado para a realização de uma partilha adequada dos dados de investigação, também contribuem para minimizar as dificuldades para a partilha dos dados em acesso aberto.

3.2. CIÊNCIA ABERTA - PRINCÍPIOS E PRÁTICAS

A CA ainda não possui um quadro teórico ou uma definição consensual que a determine e/ou a delimite (Albagli, Maciel & Abdo, 2015; Arabito & Pitrelli, 2015; Kraker et al., 2011; OCDE, 2015; Salmi, 2015), porém, uma referência amplamente usual descreve a CA como "an umbrella term encompassing a multitude of assumptions about the future of knowledge creation and dissemination" (Fecher & Friesike, 2014, p.17). Esta caracterização, intencionalmente abrangente dificulta uma percepção inicial da CA, do mesmo modo, também não se percebe um arcabouço definitivo e organizado dos seus princípios e práticas. Portanto, para selecionar e organizar os princípios e práticas da CA, tendo por finalidade verificar o que deve ser considerado para realização de uma partilha adequada dos dados de investigação, propõe-se abordar a CA sob três perspectivas: a) as origens; b) as sugestões de definição e sintetização; e, c) as políticas formuladas para a CA.

Antes de abordar as referências teóricas destacadas nas três perspectivas, é pertinente esclarecer que os termos "princípios e práticas" podem ser encontrados com outras nomenclaturas e por vezes usados indistintamente na literatura. No presente trabalho, as práticas são percebidas como uma forma de ação humana, socialmente colaborativa, realizada na tentativa do alcance dos objetivos da atividade, já os princípios, são os elementos que fundamentam a razão para que tais práticas sejam realizadas (Ilyenkov, 1982; Leont'ev, 1978). Cabe ressaltar que as referências aqui citadas, serão direcionadas para a seleção na próxima seção, resultando na organização relacional dos princípios e práticas da CA no contexto da partilha em AA. Assim, segue as três perspectivas propostas.

Na primeira perspectiva para selecionar os princípios e práticas da CA – em relação às origens -, verifica-se que a ideia de uma ciência aberta não é nova, pois um dos pilares da investigação é a premissa de que a ciência avança com base em investigações anteriores e no conhecimento acumulado. Esta premissa remonta ao século XVII, onde já havia discussões acerca da necessidade de documentar os processos de investigação e descobertas, com o intuito de explicar a ciência e minimizar o sigilo (David, 2008; Jorge & Albagli, 2017). No século passado, Bernal (1939) assinala que um dos objetivos da ciência é a divulgação e a publicação de seus resultados e em 1962 Polanyi (2000) reforçou que a comunicação aberta e a colaboração estão no centro da prática científica. Merton (1973)

descreve o comunalismo, o universalismo, o desinteresse e o ceticismo organizado¹³, como o *ethos* da ciência. Assim, podem ser percebidas no cerne destas concepções a prática da partilha, bem como os princípios da transparência e da reprodutibilidade¹⁴ da ciência, essencial no âmbito da CA.

De origem mais recente, a evolução da CA está associada à convenção de Budapeste em 2001, tornada pública a 14 de fevereiro de 2002, que constitui um marco a partir do qual ocorre um maior sentido no significado de transparência da ciência, dando início às discussões sobre o AA e encorajando novos formatos e modelos para publicação científica em periódicos abertos e auto depósitos em repositórios (BOAI, 2002; Harnard, 2005; Harnard, et al. 2008; Yiotis, 2005; Suber, 2006). Sob tal contexto e de modo progressivo, os dados de investigação foram vistos como um recurso aberto e revistas de AA passaram a exigir que os investigadores depositassem os dados relacionados aos artigos em repositórios públicos. Em consequência, a partilha dos dados de investigação em AA evoluiu em práticas, nas políticas e regulamentações no âmbito da CA (Mauthner, 2012).

Os repositórios citados nas origens da CA são ferramentas fundamentais para o movimento do AA e para a evolução da CA, uma vez que a partilha em repositórios de AA transforma significativamente a investigação científica, pois permite que os investigadores consigam acessar, analisar, avaliar, validar ou rejeitar os resultados de forma mais eficiente e transparente, possibilitando reproduzir e melhor desenvolver as investigações. Para além disso, os repositórios possibilitam realizar na prática, a transparência e a reprodutibilidade da investigação. Esta forma de partilha vai ao encontro das ideias originais da própria ciência, ou seja, favorável a uma comunicação científica aberta e colaborativa na divulgação e publicação dos seus resultados (Bernal, 1939; David, 2008; Merton, 1973; Polanyi, 1962).

¹³ O comunalismo é o princípio que define o conhecimento científico como um bem público, em oposição a um indivíduo único ou que seja restrito a um grupo de indivíduos. O universalismo refere-se à condição de que qualquer cientista pode contribuir para o avanço científico e ter a contribuição valorizada de maneira universal e impessoal. O princípio do desinteresse afirma que os cientistas devem buscar o avanço científico como uma forma de interesse coletivo e não apenas individual. O ceticismo organizado sugere que um exame cuidadoso e teste dos resultados científicos devem ser conduzidos pela comunidade antes de ser validada.

¹⁴ A transparência e a reprodutibilidade na ciência podem ser abordadas sob vários aspectos, porém de forma geral são assim explicadas: “transparency, to allow others to inspect and evaluate findings” e a “reproducibility, to verify findings by repeating a study” (Borgman, 2018).

Sob este rápido olhar sobre nas origens da CA, percebe-se que a ideia de uma ciência aberta não é nova e está imbricada com os objetivos da própria ciência, o que fundamenta e dá sentido à pergunta do título da obra de Watson (2015): “When will ‘Open Science’ become simply ‘science’?” Por outro lado, a mudança de paradigma proporcionada pela CA, talvez seja a mudança que mais afeta as práticas de condução e acesso no âmbito da investigação de forma relevante na história recente da ciência (Bartling & Friesike, 2014; Clark et al., 2021; David, 2008; Ramachandran, Bugbee, & Murphy, 2021). Esta transformação no modo de fazer e comunicar a ciência “traz de volta” princípios antigos como “transparência” e “reprodutibilidade”, atualizados em práticas de investigação viabilizadas pelas condições contemporâneas que facilitam a ampliação destes princípios em escala global, com a tecnologia e a disponibilização da produção da investigação em repositórios de AA.

A segunda perspectiva teórica para a organização dos princípios e práticas da CA, é realizada através das suas definições e propostas de sintetização. As definições para ciência aberta geralmente são compostas por termos como “conhecimento científico”, “abertura” e “partilha” no contexto de um processo. Nielsen segue esse sentido ao dizer que a “Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process” (Gezelter, 2011, para. 2). Assim, a urgência da partilha do conhecimento científico de modo aberto é necessária para garantir que o conhecimento esteja publicamente disponível, facilmente acessível e detectável para outros usarem, reutilizarem e desenvolverem, de modo a acelerar o progresso científico, transformar as descobertas e avançar a ciência para o benefício de todos (Masuzzo & Martens, 2017).

No entanto, a definição de Nielsen remete à indagação: o que significa o “conhecimento científico” que deve ser prontamente partilhado em AA? Vicente-Saez e Martinez-Fuentes (2018), com o objetivo de propor uma definição integrada para a CA, fazem uma revisão sistemática relevante, abrangente e detalhada da literatura. Os autores ratificam que o “conhecimento científico” é o termo mais encontrado nas definições da CA, atuando como seu predicado. E ainda, que os dados, produções, resultados, publicações, código, informações e ideias¹⁵ são encontradas como sinônimos de “conhecimento

¹⁵ No trabalho, estes termos aparecem como sinônimos de “conhecimento científico” na seguinte ordem: Dados, 27 vezes em 23 estudos; Publicações (científicas), 10 vezes em 10 estudos; Resultados (científicos), 9

científico”. No mesmo estudo, os termos “transparência”, “acessibilidade”, “partilha” e “desenvolvimento em colaboração” são retratados como delimitadores da CA. Assim, estes autores propõem definir a CA como “[...] transparent and accessible knowledge that is shared and developed through collaborative networks” (p.435). No estudo, a “colaboração” é destacada como um princípio fundamental para a prática da partilha em AA.

Pode-se entender que a CA e a partilha em AA, se confundem sob o conceito da abertura, de modo que a ênfase da CA recai quase sempre no AA. Porém Jorge & Albagli (2017) sugerem que a CA envolve outros aspectos a serem considerados como parte da sua dinâmica colaborativa que visa quebrar fronteiras e trabalhar concepções importantes do fazer científico. Nesse sentido, os autores definem que: “Ciência aberta não significa então apenas o acesso aberto ao conhecimento. É um movimento que envolve um processo de atualização das práticas científicas frente às transformações tecnológicas, que auxiliam na formação de redes de conhecimento e potencializam sua produção coletiva” (p.220). Assim, a partilha dos dados no contexto da CA, deve ir além de disponibilizá-los em um ambiente específico de AA, mas abarca um processo de atualização de práticas, com envolvimento tecnológico, princípios e valores fundamentais para a CA e para a sociedade. Tal contexto, que abrange pessoas e tecnologias reforça o sentido deste ambiente como sociotécnico.

A definição anterior aponta para uma mudança de paradigma de base tecnológica e social. Nesse sentido, o termo “ciência aberta”, também pode ser definido como um “Open Science brings about sociocultural and technological change, based on openness and connectivity, on how research is designed, performed, captured, and assessed.” (Seaz & Martinez-Fuentes, 2018, p.428). Essa visão da CA, compreende que as mudanças impulsionadas pela tecnologia digital direcionam para uma revolução que transforma fundamentalmente os métodos e padrões de investigação. Tal mudança permite expandir o impacto socioeconômico, remover barreiras frente ao intercâmbio de conhecimentos e alterar as práticas da criação, divulgação, armazenamento e partilha do material e do conhecimento científico (Banović, 2021; Comissão Europeia, 2016a, 2020; David, 2008).

vezes em 8 estudos; Código, 5 vezes em 4 estudos; Produções (científicas), 4 vezes em 4 estudos; Ideias, 4 vezes em 4 estudos; Informações, 3 vezes em 3 estudos.

Outras definições que podem contribuir de forma importante para o entendimento da CA, objetivado pela seleção de princípios e práticas para a partilha dos dados, são as elaboradas no âmbito do projeto *Facilitate Open Science Training for European Research* (FOSTER), a partir do contexto da sua proposta taxonômica (Pontika et al., 2015). Para FOSTER¹⁶ a “Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring Society”, ou ainda¹⁷:

Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

Nestas definições, destaca-se a importância de disponibilizar as informações científicas para uma sociedade caracterizada por tal necessidade. Assim, a prática da partilha está objetivada de modo colaborativo para a redistribuição, reprodutibilidade e reutilização¹⁸ da investigação, também através dos seus dados e métodos.

Uma recente e importante caracterização para a CA, foi realizada através da recomendação elaborada pela UNESCO (2021) sobre ciência aberta, tendo por objetivo promover um entendimento unificado sobre a CA, um consenso global. Para tanto, a UNESCO propõe a seguinte definição comum:

[...] open science is defined as an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. It comprises all scientific disciplines and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities, and it builds on the following key pillars: open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems (p.7).

A definição proposta pela UNESCO compreende uma série de termos a serem destacados como a “disponibilização do conhecimento científico” e as “práticas” que deixam o conhecimento “acessível” e “reutilizável” para todos, de modo a maximizar a

¹⁶ <https://www.fosteropenscience.eu/foster-taxonomy/open-science?page=6>

¹⁷ <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

¹⁸ A reutilização é um princípio central na ciência aberta e para a RDM, o termo quase sempre tem seu uso associado aos dados ou conjunto de dados de investigação (Palmer; Weber; Cragin, 2011).

“colaboração” e a “partilha”, abrindo os “processos” de “criação”, “avaliação” e “comunicação do conhecimento científico” que remetem ao “benefício da ciência e da sociedade”, abrangendo todas as disciplinas científicas e todas as práticas acadêmicas, tendo como pilares-chave: o conhecimento, a “infraestrutura” e a comunicação científica, bem como o “engajamento” e o diálogo com “atores sociais” e outros “sistemas de conhecimento”. Tais ideias e termos sintetizados sob esta definição, ratifica princípios e práticas já citados e introduz outros elementos importantes a serem trabalhados e explicados para a seleção proposta e para a realização da partilha em AA.

Em relação às sintetizações específicas para a CA, o projeto FOSTER propõe, define e organiza de modo conciso, através de uma taxonomia, vários componentes “abertos” dentro de uma filosofia aberta para uso na CA. São exemplos destes componentes: *Open Data*, *Open Access*, *Open Reproducible Research*, *Open Science Evaluation*, *Open Science Policies* e *Open Science Tolls* (FOSTER, n.d). Kraker et al. (2011), acrescentam mais dois instrumentos nesse contexto para a ciência aberta: *Open Source* e *Open Methodology*, este último representa a aplicação da “abertura” no fluxo de trabalho de investigação. Neste ambiente “Openness”, os componentes, *Open Access* e *Open Data*, parecem ser os instrumentos mais independentes e recorrentes da CA, pois aparecem em muitos estudos de forma individual, apesar do seu âmbito definido na CA (Tzanova, 2020).

À semelhança do projeto FOSTER, outros trabalhos propuseram sintetizar as características essenciais da ciência aberta que contemplam práticas, princípios ou conceitos sobre o assunto. Fecher e Friesike (2014), por exemplo, propõem cinco escolas de pensamentos, com suas respectivas preocupações para uma CA:

The infrastructure school (which is concerned with the technological architecture), the public school (which is concerned with the accessibility of knowledge creation), the measurement school (which is concerned with alternative impact measurement), the democratic school (which is concerned with access to knowledge) and the pragmatic school (which is concerned with collaborative research) (p.17).

Os autores desenvolvem a ideia do acesso ao conhecimento igualitário como um bem público e livremente disponível para todos, na percepção de que a promoção do engajamento dos cidadãos e a colaboração aberta são eficientes para a criação do conhecimento e do processo de descoberta. Para tanto, necessita-se do uso de infraestruturas

e ferramentas tecnológicas, redes e serviços abertos, possibilitando conexões científicas e sociais, transparência e novas possibilidades de métricas avaliativas das investigações.

A proposta de sintetização dos autores Masuzzo e Martens (2017) contribui na asserção de que o acesso aberto aos dados de investigação, aos códigos de *software*, às publicações e às revisões por pares, são os quatro pilares para posicionar a ciência aberta na prática. Nesse sentido, os autores entendem que “[...] their fundamental aim is to enhance openness in the four widely recognized thematic pillars: open research data, open software code, open access to papers, and open peer-review” (p.3).

Paul Ayris et al. (2018), em trabalho realizado pela Comissão Europeia direcionado às universidades, sintetizam e referenciam este estudo sob oito pilares para a CA:

1) O futuro da publicação acadêmica – o foco está na preparação para mudar o modelo de publicação para um acesso totalmente aberto;

2) Os princípios FAIR (*Findable, Accessible, Interoperable and Reusable*) – são projetados para ajudar a diminuir as barreiras e facilitar o acesso, a localização, a compreensão e a reutilização dos resultados e do material da investigação.

3) O ambiente virtual EOSC (*European Open Science Cloud*) – objetiva unir as infraestruturas interoperáveis existentes para permitir uma maior colaboração entre investigadores na Europa, de todos os domínios de investigação;

4) A educação e as habilidades – concentra-se em identificar as necessidades de treinamento dos investigadores e suprir lacunas no conhecimento e habilidades em relação ao envolvimento com a CA;

5) As recompensas e incentivos – é a percepção da falta de recompensa e reconhecimento pelo trabalho para tornar a produção científica aberta e acessível, abordando barreiras e práticas para encorajar os investigadores a se engajarem com a CA;

6) As métricas para a próxima geração - busca catalisar uma mudança no pensamento cultural em torno da maneira como as métricas são utilizadas na investigação, para ir além da contagem de citações e impacto de periódicos;

7) A integridade e a reprodutibilidade da investigação – é o pilar onde os investigadores agem de forma honesta, confiável, respeitosa e responsáveis por suas ações, de modo a conduzir investigações que reflitam a agenda da CA; e

8) A ciência cidadã – aborda uma mudança na forma como a investigação é conduzida - não limitada à acadêmicos, mas na colaboração de grupos e de toda a sociedade.

Bezjak et.al (2018), no trabalho orientado pelos princípios e práticas da CA, sintetizam-na de maneira a inserir características importantes para seu uso e contexto:

Open Science is about increased rigour, accountability, and reproducibility for research. It is based on the principles of inclusion, fairness, equity, and sharing, and ultimately seeks to change the way research is done, who is involved and how it is valued. It aims to make research more open to participation, review/refutation, improvement and (re)use for the world to benefit [...]. Open Science encompasses a variety of practices, usually including areas like open access to publications, open research data, open source software/tools, open workflows, citizen science, open educational resources, and alternative methods for research evaluation including open peer review (p.12).

Os autores tratam como princípios: a inclusão, a equidade, a justiça e a partilha. Tais princípios resultam no aumento do rigor, da responsabilidade e da reprodutibilidade da investigação, mas também reconhecem a influência da CA na forma e na valorização da investigação, de modo a propiciar uma abertura participativa para a revisão, refutação, melhoria e reutilização da investigação em benefícios de todos. Os autores ainda destacam as práticas que incluem a partilha dos dados em AA, bem como as publicações, *softwares* e ferramentas de código aberto, métodos para a avaliação da investigação e revisão por pares. Esta sintetização demonstra a amplitude com que os princípios e práticas da CA podem ser trabalhados no contexto da partilha dos dados em acesso aberto.

Enquanto na primeira perspectiva para a organização dos princípios e práticas da CA, os princípios da “transparência” e da “reprodutibilidade” foram destacados para a necessidade da atualização das práticas de investigação e partilha em repositórios de AA, nesta segunda perspectiva as referências que reportam aos conceitos e as propostas de sintetizações da CA, por sua diversidade e amplitude, permitem selecionar outros pontos de interesse para este estudo, tais como: a) a ênfase na comunicação e no “conhecimento científico”; b) o destaque no processo e na colaboração; c) o uso da tecnologia nas atualizações das práticas; d) a mudança de paradigma de base tecnológica e sociocultural;

e) o engajamento da sociedade e os benefícios sociais; f) a filosofia e práticas abertas - “openness”, da ciência; e, g) as características para avaliação e avanço da ciência, como a revisão, refutação, melhoria e reutilização dos dados e do material da investigação.

A terceira e última perspectiva para a selecionar e organizar os princípios e práticas da CA, refere-se às políticas formuladas para a ciência aberta. Tais políticas podem ser entendidas como ações e estratégias destinadas a promover e implementar os princípios e as práticas da CA. Estas políticas têm sido formuladas como alvo de atenção de governos, instituições de investigação, agências de fomento e financiadoras da ciência em todo o mundo, tais como: os *Research Councils UK* (RCUK) e o *Wellcome Trust* no Reino Unido, a *White House’s Office of Science and Technology Policy* (OSTP), o *Department of Defense* (DoD), *Department of Energy*, a *National Science Foundation* (NSF) e a *National Institutes of Health* (NIH) nos Estados Unidos, a *German Research Foundation* (DFG) na Alemanha e a Comissão Europeia (Borgman, 2018; Elliott & Resnik, 2019; Jones, 2011 OCDE, 2021; Royal Society 2012; Shearer 2015). Estas instituições, sob o fundamento da CA, passaram a recomendar ou exigir através das políticas formuladas, a partilha em AA como requisito fundamental para as suas atividades e acesso aos seus financiamentos.

Em relação as políticas nacionais de AA, o *Open Access Infrastructure for Research in Europe* (OpenAIRE, 2021) apresenta exemplos de diversos países que podem ajudar na observação dos princípios e práticas da CA. São aqui destacadas as políticas da Eslováquia, Holanda, Bulgária, Finlândia, Irlanda e Portugal.

Na Eslováquia a estratégia nacional para a CA visa melhorar a disponibilidade dos resultados da ciência, através dos seguintes fundamentos: a) o acesso aberto a publicações com financiamento público; b) o acesso aberto aos dados de investigação; c) a infraestrutura técnica para CA; d) o financiamento da CA; e) a proteção de direitos de propriedade intelectual; f) o uso de ferramentas abertas de tecnologia da informação; g) a educação em CA; h) a avaliação de unidades de I&D com princípios da CA; e, i) o apoio à Ciência Cidadã.

Na Holanda, o plano nacional de ciência aberta é direcionado para acelerar as seguintes práticas: a) a publicação total em acesso aberto; b) a reutilização de dados de investigação; c) a avaliação para o reconhecimento e recompensas; d) o encorajamento e

apoio à CA; f) a profissionalização da gestão de dados; e, a g) ciência cidadã. Ressalta-se que a política de CA dos financiadores na Holanda, exigem a publicação e a disponibilização dos dados em AA com ênfase em possível uso futuro.

O plano nacional da CA da Bulgária descreve compromissos e pontos de ação no que diz respeito ao desenvolvimento e a sustentabilidade da CA, porém dois pontos são destacados: 1) as investigações financiadas pelo Estado, na qual o material resultante deve ser inserido nos bancos de dados de todas as instituições de investigação ou conectados em um portal conjunto para investigações; e, 2) a criação de um serviço para armazenamento em longo prazo, para as publicações e para que os dados de investigação possam ser lidos e utilizados para o acesso no presente e no futuro.

Na Finlândia, os objetivos da política e da declaração para uma ciência aberta incluem o reforço na infraestrutura da investigação e a salvaguarda da abertura da investigação e da ciência. O documento também define como a abertura se tornará parte da vida diária dos investigadores através de quatro objetivos: 1) a cultura de investigação; 2) o acesso aberto as publicações; 3) o acesso aberto aos dados e métodos de investigação; e, 4) a educação aberta e recursos educacionais promovidos pela comunidade de investigação.

Em destaque na Irlanda está o objetivo da transição para um ambiente de investigação aberto. O plano irlandês prevê ações para permitir o AA às publicações, possibilitar dados FAIR, desenvolver habilidades e competências, criar incentivos, recompensas e infraestruturas para o acesso e a preservação dos materiais da investigação. Os princípios nacionais de AA ainda asseveram que:

a) Os artigos de periódicos revisados por pares e outros resultados de investigação com financiamento público, devem ser depositados em repositório de AA, os tornando disponíveis, acessíveis e reutilizáveis o mais rápido possível e de forma contínua;

b) Cada investigador deve ter direitos de depósito em um repositório de AA;

c) Os autores devem depositar as pós-impresões mais os metadados dos artigos;

d) O depósito deve ser feito, de preferência no momento da aceitação para publicação, e o mais tardar na data da publicação formal; e

e) Os metadados devem incluir os dados descritivos e devem estar conforme as normas e acordos nacionais e internacionais para recolha e interoperabilidade.

Em Portugal, embora até à data deste trabalho não se tenha implementado uma política nacional de CA, o Governo português declara a CA como uma prioridade. O grupo de trabalho interministerial criado pelo Ministério da Ciência, Tecnologia e Ensino Superior, também redigiu um conjunto de recomendações, dirigidas a distintas partes interessadas, para a implementação de uma política nacional de CA, nas dimensões i) do acesso aberto e dos dados abertos; ii) das infraestruturas de acesso e preservação digital; iii) da avaliação científica; e, iv) da responsabilidade social científica¹⁹.

O *site* nacional português²⁰ para a CA, reúne informações, iniciativas, conteúdos formativos, resoluções e documentos relevantes que assinalam como itens base para a ciência aberta: a) o acesso aberto; b) os dados abertos; c) a investigação/ inovação aberta; d) as redes abertas de ciência; e, e) a ciência cidadã. No mesmo sentido, indica três ações estruturantes: 1) a transparência nas práticas, metodologia, observação e recolha de dados; 2) a disponibilização pública e a reutilização de dados, o acesso público e transparente na comunicação científica; e, 3) a utilização de ferramentas baseadas na *Web* para facilitar a colaboração científica. E ainda lista oito vantagens para a abertura da ciência:

- 1) O aumento da eficiência na investigação;
- 2) O aumento do conhecimento do processo de trabalho científico;
- 3) A promoção do rigor académico e aumento da qualidade da investigação;
- 4) A aceleração da criação de novos temas de investigação;
- 5) A promoção do envolvimento da sociedade e da cultura/literacia científica;
- 6) O aumento do impacto económico e social da ciência;
- 7) A valorização da propriedade intelectual; e
- 8) A promoção do retorno científico para as instituições.

¹⁹ https://www.ciencia-aberta.pt/_files/ugd/a8bd7c_3274046fc8ce42c78db2ec1707c0a0fd.pdf

²⁰ <https://www.ciencia-aberta.pt/>

Além dos governos e órgãos nacionais, instituições como a OCDE formulam políticas de interesse público tendo por base a CA. Tais políticas podem funcionar como um agregador de princípios e práticas, útil para a seleção aqui proposta. Assim, o documento da OCDE (2021)²¹, original de 2006, direcionado à partilha dos dados de investigação com financiamento público em AA, reafirma e destaca a relevância dos seguintes princípios:

- a) Abertura – acesso em termos iguais para a comunidade de investigação;
- b) Flexibilidade – adaptação às mudanças rápidas na tecnologia da informação;
- c) Transparência – dos dados, organizações e documentação;
- d) Conformidade jurídica – a respeito dos interesses das partes interessadas;
- e) Proteção da propriedade intelectual – direitos autorais, intelectuais e de propriedade;
- f) Responsabilidade formal – formulação de regras e regulamentos para as várias partes envolvidas;
- g) Profissionalismo – uso de valores profissionais dos códigos de conduta nas comunidades científicas;
- h) Interoperabilidade – atenção em termos de tecnologia e semântica;
- i) Qualidade – garantia dos padrões para os dados;
- j) Segurança – garantia da integridade e salvaguarda dos dados;
- k) Eficiência - acesso e partilha para melhoria geral da investigação;
- l) Responsabilidade – na avaliação do valor e do acesso aos dados; e
- m) Sustentabilidade – garantia do acesso em longo prazo.

A OCDE justifica que os princípios da CA devem ser realizados com a prática da partilha dos dados em AA, por reconhecer o rápido crescimento dos dados e os novos paradigmas da inovação sob uma ciência intensiva em dados em um contexto que está a

²¹ No contexto da pandemia global COVID-19, a OCDE atualizou as políticas que permitem a partilha de dados de investigação, considerando-as como importantes para ajudar a acelerar o combate à doença. A recomendação atualiza as orientações de 2006 sobre o aprimoramento do acesso aos dados de investigação.

transformar o panorama da política científica e tecnológica. Ainda segundo a OCDE, objetos digitais relevantes para investigação como - *software*, algoritmos, fluxos de trabalho científicos, códigos e o ambiente computacional -, tornam-se um bem público que pode criar valor para a sociedade quando tais objetos são usados para gerar resultados juntamente com os dados e metadados, disponíveis para acesso rápido e eficiente, sendo essenciais para:

- a) Viabilizar a reprodutibilidade dos resultados científicos;
- b) Reduzir a duplicação do esforço por meio da reutilização;
- c) Criar oportunidades para novos *insights* científicos;
- d) Possibilitar as investigações interdisciplinares;
- e) Acelerar os avanços da ciência e da tecnologia;
- f) Promover a descoberta científica;
- g) Permitir o desenvolvimento da inteligência artificial e aplicações emergentes;
- h) Viabilizar a inovação, progresso tecnológico e o crescimento econômico;
- i) Reutilizar e valorizar os dados;
- j) Aumentar o bem-estar dos indivíduos e da sociedade em geral; e
- k) Permitir que os investigadores resolvam desafios globais de forma eficiente.

Além das justificativas, as recomendações da OCDE apontam para as seguintes preocupações e pretensões:

- a) Melhorar o acesso aos dados de investigação e torná-los FAIR;
- b) Partilhar os conjuntos de dados menores da “cauda longa”;
- c) Fornecer acesso seguro, justo, inclusivo, legal, ético e responsável;
- d) Evitar a obsolescência dos dados e *software*, com documentação de metadados, manutenção ativa e atualizada do formato de dados e do design de *software*, bem como indicar os portais e repositórios onde os objetos são armazenados;

- e) Realizar medidas para proteger interesses privados, públicos e comunitários, incluindo a segurança nacional, os direitos de propriedade intelectual, privacidade, os dados pessoais, o meio ambiente, os recursos naturais essenciais, incluindo água e minerais, produtos e espécies ameaçadas;
- f) Priorizar o fornecimento do acesso a objetos digitais relevantes para a investigação, dado o consumo de muitos recursos e o volume exponencialmente crescente dos dados e de outros objetos digitais;
- g) Valorizar habilidades especializadas e o capital humano, para aumentar a confiança e aproveitar todos os benefícios potenciais de um maior acesso, partilha e reutilização de dados e outros objetos digitais relevantes;
- h) Remover barreiras significativas ao fornecimento de acesso aos dados sensíveis, de forma restrita e em ambientes seguros e protegidos; e
- i) Promover a interoperabilidade dos quadros jurídicos internacionais que garantam uma proteção consistente contra a utilização indevida dos dados.

Na recomendação elaborada pela UNESCO (2021) sobre ciência aberta, além da definição já citada, são elencados os seguintes princípios orientadores:

- a) Transparência, escrutínio, crítica e reprodutibilidade – através de promoção de uma maior abertura em todas as etapas do empreendimento científico;
- b) Igualdade de oportunidades – onde todos os cientistas e atores devem ter as mesmas oportunidades de acesso, contribuição benefícios da ciência aberta;
- c) Responsabilidade, respeito e prestação de contas – de forma pública para todos os atores da CA, em relação aos princípios éticos relativos à investigação;
- d) Colaboração, participação e inclusão - em todos os níveis do processo científico juntamente com a participação plena e efetiva dos atores sociais;
- e) Flexibilidade - decorrente da diversidade dos sistemas científicos e da natureza em constante evolução das tecnologias de informação e comunicação; e
- f) Sustentabilidade - com práticas, serviços, infraestruturas e modelos de financiamento que garantem o acesso de longo prazo, na medida do possível.

As políticas formuladas para a CA são bem detalhadas e parecem ser capazes de justificar o uso da partilha dos dados em AA, bem como orientar o modo de fazê-lo. Portanto, dentre os pormenores e as generalizações, pode-se destacar alguns princípios e práticas oriundas das políticas formuladas em torno da CA como, por exemplo: a) a proteção de direitos de propriedade intelectual; b) o depósito de metadados; c) o reconhecimento e as recompensas; d) a profissionalização da gestão de dados; e) a garantia do acesso em longo prazo; f) as infraestruturas para acesso e preservação da produção científica; g) a redução da duplicidade da investigação; h) os novos *insights* científicos a serem gerados; e, i) as questões sobre interoperabilidade, flexibilidade e responsabilidades decorrentes.

Nesta seção, muitos foram os princípios e as práticas importantes citados e a serem considerados para uma partilha adequada dos dados de investigação. Neste sentido, tendo por contexto a partilha em AA, a seção que segue seleciona e organiza os temas e a relação entre os princípios e práticas da CA, através dos vocábulos envolvidos nas várias citações teóricas em relação: as origens; as sugestões de definição e sintetização e as políticas formuladas para a CA, resultando em uma proposta de organização relacional temática.

3.3. UMA ORGANIZAÇÃO RELACIONAL TEMÁTICA

No entendimento de que a CA “can be defined as a grouping of principles and practices” (Bezjak et al., 2018, p.8) se encontram as referências até aqui trabalhadas. No entanto, o conjunto amplo de itens, pontos, questões, pormenores e vieses dos princípios e práticas espalhados nas referências, podem ser estudados de diversas maneiras sob o “guarda-chuva” da CA. Com isso, é necessário não incorrer no erro em referenciar uma “chuva” de termos, conceitos e palavras sem um sentido organizado (Figura 4). Tais termos e palavras, embora importantes, podem ser dispersos e carentes de sentido quando não atribuídos a um contexto teórico ou, quando não relacionados entre si e com a referência que os originou. Portanto, os princípios e práticas da CA necessitam de uma organização relacional temática.

A organização proposta nesta seção, possibilita que os temas espalhados pelas citações realizadas neste capítulo sejam organizados e relacionados, facilitando assim a verificação do que deve ser considerado para realizar uma partilha adequada dos dados de investigação, tendo em vista o ambiente da universidade e o contexto do acesso aberto.

Tabela 9

Termos e palavras oriundas dos princípios e práticas da ciência aberta

Lista de termos e palavras que envolvem os princípios e práticas da ciência aberta e a partilha em acesso aberto			
Acesso	Direitos	Melhoria	Público
Algoritmos	Disponibilidade	Métricas	Qualidade
Ambiente computacional	Documentação	Mudança	Recolha de dados
Análise	Duplicidade	Obsolescência	Recompensas
Apoio	Econômico	Open Access	Redes
Armazenamento	Educação	Open access to papers	Redistribuição
Atualização	Eficiência	Open access to publications	Refutação
Avaliação	Encorajamento	Open Data	Regras
Avanço da ciência	Engajamento	Open educational resources	Regulamentos
Barreiras	Envolvimento	Open Methodology	Rejeição
Bem-estar social	Equidade	Open peer-review	Repositórios
Benefício social	Estratégias	Open Reproducible Research	Reprodutibilidade
Capital humano	Ética	Open Science Evaluation	Responsabilidade
Cauda longa	Ferramentas	Open Science Policies	Respeito
Ceticismo organizado	Financiamento	Open Science Tolls	Resultados
Ciência cidadã	Flexibilidade	Open software code	Reutilização
Códigos	Futuro	Open Source	Revisão
Colaboração	Gratuidade	Open-source software/tools	Rigor
Competências	Habilidades	Open workflows	Salvaguarda
Comunalismo	Igualdade	Padrões	Segurança
Comunidade científica	Impacto	Paradigmas	Sociocultural
Condução	Incentivos	Políticas	Sociotécnico
Conectividade	Inclusão	Pragmática	Software
Conformidade	Infraestrutura	Pré e pós-prints	Sustentabilidade
Conhecimento	Inovação	Preservação	Técnico
Continuidade	Insights	Priorização	Temas novos
Contribuição	Integridade	Processo de trabalho	Transição
Dados FAIR	Interdisciplina	Profissionalismo	Transparência
Dados sensíveis	Interesses	Profissionalização	Universalismo
Democrática	Interoperabilidade	Progresso científico	Urgência
Depósitos	Justiça	Projeto	Usabilidade
Desafios globais	Legalidade	Propriedade intelectual	Validação
Descoberta	Literacia	Proteção	Valorização
Desinteresse	Longo prazo	Publicação	Volume

É interessante observar a força e a importância que os termos e as palavras selecionadas possuem para a CA no contexto da partilha em AA. Por tal relevância, parece ser possível desenvolver abordagens distintas sobre cada um dos itens listados. No entanto, embora pareça ser exequível o seu uso individualizado, para uma organização relacional, faz-se necessário realizar um agrupamento dos itens em aspectos temáticos.

A proposição de um agrupamento deste tipo é de difícil realização, pois existem várias temáticas espalhadas pelos princípios e práticas sob os quais os termos e as palavras foram originadas e referenciadas. Assim, para agrupá-los em aspectos é preciso perceber o propósito original que os representa, observar os significados relacionados que os qualifica e atentar ao objeto da referência. Portanto, sob esses cuidados, os princípios e práticas da CA sob o contexto da partilha em AA foram agrupados em oito aspectos temáticos, quais sejam: 1) científicos; 2) colaborativos; 3) legais; 4) *openness* – abertos; 5) pessoais; 6) sociais; 7) sustentáveis; e, 8) técnicos e materiais.

Como resultado, a tabela abaixo apresenta o agrupamento constituído a partir dos termos e palavras selecionados e listados na tabela anterior. Onde, cada aspecto temático agrupa princípios ou práticas individualmente referenciados ou, entre eles associados para ganho de significado. Assim, na construção da tabela, os princípios e as práticas foram agrupados de acordo com os seus propósitos principais, observados na referência como mais adequados ao aspecto temático atribuído. Igualmente, o aspecto temático foi concebido tendo por base os propósitos centrais dos princípios e práticas, ou seja, ambos se constituem.

Tabela 10

Princípios e práticas da ciência aberta agrupados por aspectos temáticos

Aspectos temáticos	Princípios e práticas da ciência aberta para a partilha em acesso aberto
Científicos	Acesso ao conhecimento; acesso aos resultados; ceticismo organizado; avaliação dos resultados e da investigação; avanço da ciência; comunicação da ciência; condução da investigação; disponibilização da produção digital; divulgação dos resultados e do conhecimento; documentação dos processos de investigação; eficiência da ciência; equidade; exame dos resultados; explicação da ciência; inovação; igualdade; integridade dos dados e da investigação; interdisciplinaridade; justiça; descobertas científicas; publicação dos resultados; qualidade dos dados e da investigação; recolha dos dados; redistribuição dos dados, métodos e material subjacentes; redução da duplicidade; refutação da investigação; rejeição dos resultados; reprodutibilidade e reutilização dos dados, métodos e material subjacentes; revisão por pares; rigor acadêmico; transparência e validação dos resultados.

Colaborativos	Comunidade científica em rede; conexões acadêmicas e sociais; contribuição e colaboração de “outros”; desenvolvimento por redes; desinteresse individual; envolvimento via ferramentas <i>web</i> ; estrutura partilhada; comunicação colaborativa; instituições de investigação conectadas e participação.
Legais	Acesso legal e ético; códigos de conduta; conformidade jurídica; controle autoral; dados pessoais; direito de reconhecimento e citação; direitos autorais, intelectuais e de propriedade; ética; limite para fornecimento de dados sensíveis e sigilosos; privacidade e segurança dos dados e material; proteção de interesses privados, públicos, comunitários e meio ambiente; responsabilidade formal; emprego de regras e regulamentos; segurança nacional; partes interessadas e uso indevido dos dados.
Openness (abertos)	Acesso justo, online e livre; atualização para práticas abertas; ausência de barreiras; cauda longa da ciência; comunicação aberta; disponibilidade digital; FAIR; financiamento público; flexibilidade; gratuidade; livre acesso ao conhecimento, métodos e padrões; métricas de impacto; mínimas restrições e custo; mudança de paradigma; open access to papers; open access to publications; open data; open educational resources; open methodology; open peer-review; open reproducible research; open research data; open science evaluation; open science policies; open science tolls; open software code; open source software/tools; open workflow; recursos abertos; transição e urgência.
Pessoais	Acesso responsável; apoio; capacitação; capital humano; competência; educação; encorajamento; habilidades; incentivos; interação sociotécnica; literacia científica; profissionalismo; profissionalização; recompensas; reconhecimento; responsabilidade e valorização.
Sociais	Acesso inclusivo; bem-público; bem-estar social; benefício social; ciência cidadã; comunalismo; conhecimento igualitário; disponibilidade para todos; engajamento dos cidadãos, sociedade e cultura; impacto e melhorias socioeconômico, sociocultural, territorial e ambiental; interesse coletivo; prevenção, mitigação e soluções para os desafios e crises globais; redução da desigualdade informacional; responsabilidade social científica; sociedade inquiridora e universalismo.
Sustentáveis	Armazenamento em longo prazo; continuidade; documentação, planejamento, estratégia; garantia de acesso presente e futuro; manutenção; obsolescência; possível uso futuro; preservação dos materiais; projeção e salvaguarda da abertura.
Técnicos e Materiais	Acesso seguro, rápido e eficiente; armazenamento da produção científica; depósito em repositórios digitais; desenvolvimento de ambiente computacional, infraestruturas técnicas de redes, ferramentas de TI e serviços; guarda de objetos digitais relevantes para investigação – dados, algoritmos, códigos de software, publicações, fluxos de trabalho científicos, metadados, documentações, pré e pós prints; interoperabilidade; padronização dos materiais; priorização de acesso; reforço na infraestrutura da investigação; segurança e limite tecnológico; transformação tecnológica; usabilidade; valorização e volume.

Cabe observar, que embora tenha-se procurado atribuir um item por aspecto temático, de forma a evitar a repetição por aspectos diferentes, tendo em vista a abrangência do assunto, os itens não estão limitados e podem ser utilizados em mais de um aspecto. No entanto, os oito aspectos temáticos propostos que agrupam os princípios e práticas da CA no contexto da partilha em AA necessitam ser explicados, relacionados e contextualizados à luz das referências sob a qual foram selecionados, é o que segue.

O aspecto científico agrupa os princípios e as práticas que são vinculadas à transparência da ciência, com forte fundamento na partilha. Este aspecto toma por base a premissa de que a ciência avança com suporte em investigações anteriores e no conhecimento acumulado. Neste sentido, as referências apontam para o progresso da área científica através das inovações e das descobertas. Assim, é necessário que o processo da condução da investigação, a começar antes mesmo da recolha dos dados, seja direcionado à partilha dos produtos da investigação em AA. Ou seja, a publicação dos resultados da produção digital – dados, métodos, documentação e demais materiais. Igualmente, a redistribuição destes materiais em AA também permite que a ciência avance, e sua reutilização colabore com o rigor e com a eficiência científica, no sentido de que as ações de validação (ou rejeição) da investigação sejam facilitados pela partilha em AA como, por exemplo, a revisão por pares. Outra característica deste aspecto está na igualdade de oportunidades científicas para os investigadores, independentemente do local onde estejam, proporcionando o acesso justo em AA, permitindo também garantir a equidade entre os investigadores. Neste aspecto temático, são citados como vantagens para o avanço da ciência: a transparência e a comunicação científica, a possibilidade da interdisciplinaridade, da redução da duplicidade na investigação e de uma melhor explicação da ciência.

O aspecto colaborativo parte do princípio de que a ciência aberta está ancorada na colaboração. Todos os aspectos propostos dependem e são constituídos por valores colaborativos para que a CA e a partilha em AA seja possível. Assim, este aspecto temático está orientado em princípios e práticas que tornam o processo científico acessível e aberto em redes de conhecimento e comunicação. O desenvolvimento por redes é referido como central na colaboração para a partilha da produção coletiva em AA. Tal colaboração envolve necessariamente um esforço comum e apreciativo de vários atores, fluxos, ferramentas e infraestruturas para disponibilização e acesso ao material da investigação. Sendo assim, o aspecto colaborativo pode ser desenvolvido de modo que os investigadores e outros atores possam participar de forma inclusiva, com conexões cooperativas e através de redes entre os investigadores, entre os investigadores e as comunidades acadêmicas e entre ambos e a sociedade.

O aspecto legal abrange o arcabouço referencial de princípios e práticas que envolvem as leis de privacidade e segurança, licenças, contratos de confidencialidade, normas, códigos de ética e de condutas e outros mecanismos semelhantes que direcionam às proteções necessárias para que os investigadores possam partilhar em AA. Neste sentido, é reconhecido que a partilha em AA pode proporcionar vulnerabilidades quando não observada a conformidade jurídica para os direitos existentes. Assim, as referências e o sentido de uso do aspecto legal seguem um caminho para minimizar as vulnerabilidades e proporcionar segurança para o contexto da investigação e dos seus atores. Isto pode ocorrer através de ações como por exemplo: as que viabilizam a proteção da propriedade intelectual, os direitos dos autores e investigadores às suas produções científicas, ou de serem reconhecidos e citados. Na mesma direção, também são alvos da observação deste aspecto para partilha em AA o tratamento e os limites para o acesso ao material da investigação, como os dados sensíveis e sigilosos e as informações pessoais, privadas ou públicas que possam incorrer em danos particulares, econômicos e sociais.

O aspecto *openness* agrupa os princípios e as práticas referentes à filosofia aberta, fundamentados no livre acesso ao conhecimento para a partilha em AA. Este aspecto abrange a necessidade do acesso *online*, gratuito e com ausência ou minimização de barreiras para disponibilização do material da investigação, de preferência, com o tratamento dos dados nos princípios FAIR, mas também engloba a atualização de práticas necessárias para que o investigador e a investigação possam se adaptar ao paradigma aberto. Por isso, remete à urgência e a necessidade de flexibilidade e de transição para o paradigma em questão, tendo em vista a tendência da velocidade e os grandes volumes com que os materiais de investigação digitais são produzidos. O aspecto abarca também a exigência da partilha proveniente das instituições que financiam a investigação, com destaque para o material produzido na “cauda longa da ciência”. Considera-se ainda que uma série de novas práticas da ciência, mas também outras já conhecidas, encontram-se perante a filosofia aberta identificadas sob o termo *open*: são métodos, métricas, publicações, dados, recursos educacionais, metodologias, revisões, avaliações e políticas, entre outros que caracterizam este aspecto no contexto da partilha em AA.

O aspecto pessoal agrupa o capital humano como alvo de atenção referencial. Este aspecto evidencia a importância das pessoas no contexto da partilha em AA, as referências destacam a necessidade de apoio e capacitação para desenvolver e reforçar as habilidades e as competências necessárias para a partilha, como a preparação para uma interação sociotécnica. Mas também envolve fatores que remetem à valorização pessoal, principalmente em forma de reconhecimentos, recompensas e incentivos sob o paradigma da CA. No aspecto pessoal, o ambiente de trabalho é destacado pela necessidade de profissionalismo e profissionalização na preparação e execução da partilha em AA, que pode ser realizado por intermédio do desenvolvimento de uma estrutura definidora de responsabilidades e divisão do trabalho, com objetivo de formar um ambiente adequado para que os produtos da investigação sejam partilhados em AA.

O aspecto social é direcionado pelas referências que tratam a partilha em AA como um bem público, motivados pelo financiamento público e pela responsabilidade social científica. O uso destes princípios e práticas devem resultar no bem-estar social, em benefícios para a sociedade e soluções para interesses coletivos. Sob a referência da disponibilidade e acesso para todos, o aspecto social pode permitir reduzir a desigualdade informacional e proporcionar uma inclusão igualitária. Neste aspecto, se espera que uma sociedade inquiridora engaje cidadãos na participação da ciência, para que tal envolvimento possa resultar em soluções no enfrentamento dos problemas sociais, melhorias socioeconômicas, socioculturais, territoriais e ambientais.

O aspecto sustentável pode ser o menos perceptível entre os aspectos aqui apresentados, devido à possibilidade do imediatismo e da urgência das ações a serem realizadas no contexto da investigação. Em consequência, corre-se o risco de não distinguir ações em curto ou longo prazo para partilha em AA. Por exemplo, a prática de armazenamento pode ser cumprida com o depósito do material em um repositório, porém as referências preocupadas com a sustentabilidade remetem a um esforço para observar que esta prática deve ser preparada sob a ideia do “longo prazo”. No mesmo sentido, a documentação a ser produzida deve ser planeada de modo a servir para interpretação e uso do material no futuro, e ainda, que o ato da partilha, manutenção e atualização das estruturas e infraestruturas devem ser realizadas de forma contínua, garantindo assim o acesso presente

e futuro para salvaguardar a abertura do material partilhado. Este aspecto reconhece que a obsolescência faz parte do processo natural que envolve a partilha em AA, por isso deve ser alvo de atenção e de ações para possibilitar a sustentabilidade dos produtos partilhados.

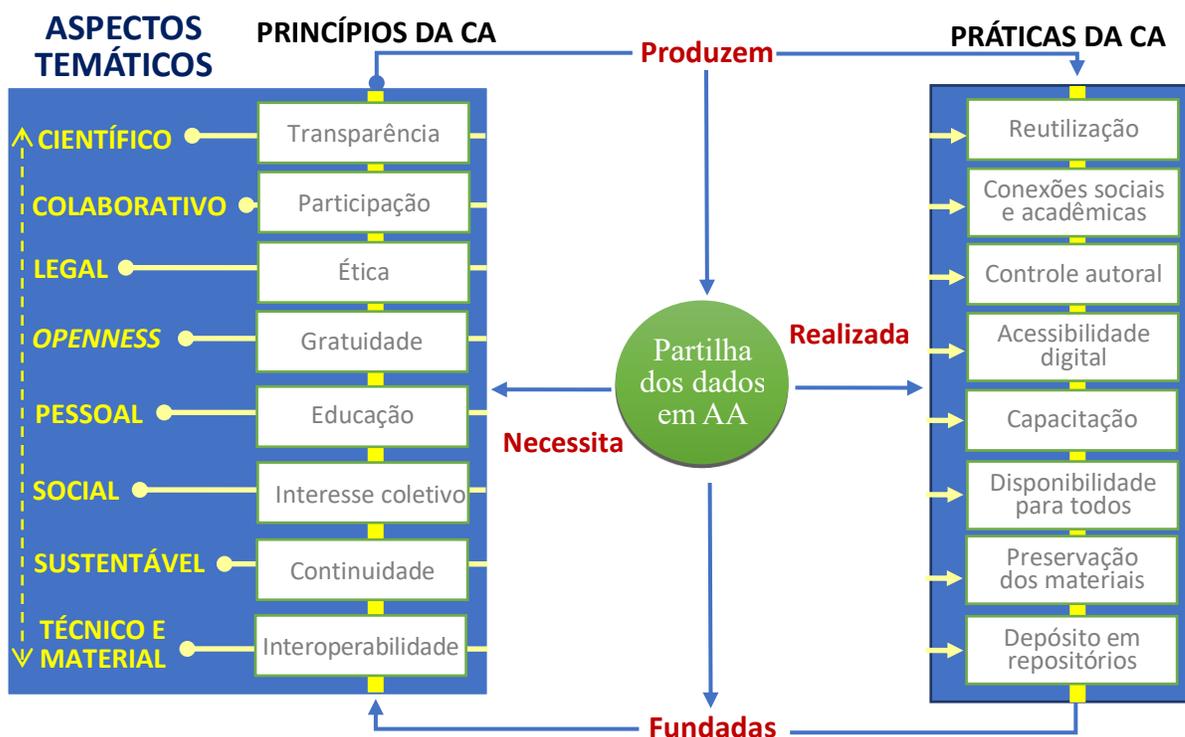
O aspecto técnico e material agrupa os princípios e práticas que possibilitam a disponibilização e o acesso seguro, rápido e eficiente para a partilha em acesso aberto. A partir deste aspecto, ocorrem as condições infraestruturais para a guarda da produção científica em repositórios digitais. São ferramentas e serviços interligados, permeados por um ambiente tecnológico e computacional para guarda de dados, algoritmos, códigos de *software*, publicações, fluxos de trabalho científicos, metadados, documentações e outros materiais importantes para a investigação. O aspecto técnico e material exemplifica bem o envolvimento dos outros aspectos temáticos e seus inter-relacionamentos. Por exemplo, na prática do aspecto material, as restrições, níveis de acesso, limites e tratamento dos dados podem ser tecnicamente estabelecidos, isto proporciona, em parte, o cumprimento dos preceitos de segurança e proteção de direitos característicos no aspecto legal. A usabilidade das ferramentas, também desenvolvidas no aspecto técnico e material, possibilita a facilidade de uso para as pessoas, contribuindo com o aspecto pessoal. No aspecto técnico, a interoperabilidade dos materiais precisa ser desenvolvida para que o aspecto colaborativo possa ser maximizado. E ainda, o reforço e o acompanhamento da infraestrutura técnica e material da investigação para a partilha em AA remete ao aspecto sustentável, pois na realização de intervenções e manutenções, consegue-se suportar ou limitar o volume do material da investigação a ser partilhado.

Conforme exemplificado no parágrafo anterior, várias são as conexões possíveis e necessárias entre os oito aspectos temáticos, tendo em vista a conjunção de propósitos teóricos semelhantes agrupados sob uma mesma finalidade. Assim, para considerar os princípios e as práticas que podem resultar em uma partilha adequada dos dados de investigação, pode-se perceber que estes aspectos temáticos dependem e se relacionam uns com os outros e com a partilha em AA. Portanto, apesar dos aspectos temáticos poderem ter outras valias, para os objetivos desta investigação, tais aspectos podem organizar de forma lógica e interrelacionada os princípios e as práticas da CA, objetivada para a partilha dos dados em acesso aberto, conforme apresentado na figura 5.

É no sentido das relações interdependentes que o exemplo abaixo propõe uma organização relacional temática, exemplificando o uso dos aspectos temáticos e consequentemente, dos princípios e práticas em interação com uma partilha dos dados em AA. Nesse entendimento, propõe-se uma relação recíproca entre os aspectos temáticos, seus constituintes e uma partilha de dados em AA, conforme exemplificado na figura que segue.

Figura 5

Exemplo de uma organização relacional temática dos princípios e práticas da ciência aberta



A organização relacional temática exemplificada resulta no mínimo em duas conclusões: a primeira está na percepção de que uma partilha dos dados em AA não se completa de forma adequada sem um propósito, que seja lógico e ancorado nos princípios e práticas da CA. A segunda é de que os aspectos temáticos constituídos pelos princípios e práticas da CA, necessitam estar relacionados entre si para realizar adequadamente uma partilha dos dados em AA. Assim, a primeira conclusão evidencia a necessidade de um propósito racionalizado e a segunda destaca a necessidade dos relacionamentos. Ambas podem ser vistas para contrapor o risco da “partilha pela partilha” e das práticas isoladas. Segue a explicação da figura acima de acordo com estas duas conclusões.

A primeira conclusão pode ser exemplificada e descrita no seguinte sentido: 1) Os princípios da CA são elementos que **produzem** práticas e fornecem os insumos necessários para que uma partilha em AA seja realizada; 2) Uma partilha em AA para ser adequada, **necessita** ser **fundada** nos princípios da CA e **realizada** no correto cumprimento das práticas da CA; 3) As práticas da CA devem ser **fundadas** nos princípios da CA para que assim cumpram a finalidade pelas quais foram **produzidas** e que devem ser **realizadas**.

A segunda conclusão pode ser percebida e exemplificada através do princípio da transparência e da prática da reutilização, nos seguintes termos: para a reutilização dos dados e do material partilhado em acesso aberto (aspecto científico – princípio da transparência), é necessário o depósito em um repositório digital (aspecto técnico e material – princípio da infraestrutura) que possibilite a preservação e a reutilização presente e futura de tal material (aspecto sustentável – princípio da continuidade), estando disponível e acessível para todos os que precisarem (aspecto social – princípio do benefício social), permitindo o uso para várias finalidades, como por exemplo, em conexões acadêmicas e sociais (aspecto colaborativo – princípio da participação). No entanto, para que isto aconteça corretamente é preciso garantir os direitos dos atores envolvidos, como, por exemplo, o controle autoral (aspecto legal – princípio ético) e a capacitação dos utilizadores (aspecto pessoal – princípio da educação). Assim, pelo relacionamento de todos estes aspectos, pode-se assegurar a partilha em acesso aberto para a reutilização do material da investigação.

No decorrer desta seção, realizou-se a seleção dos termos e palavras das referências que fundamentaram os princípios e práticas da CA e da partilha em AA, abordados nas seções anteriores. Em um segundo momento, tais termos e palavras foram agrupados em aspectos temáticos. Por fim, dos aspectos temáticos derivaram uma organização para a partilha dos dados em AA, enfatizando o seu uso através de um propósito racionalizado e na necessidade dos relacionamentos. Assim, o caminho percorrido resultou em uma organização relacional temática orientada pelos princípios e práticas da CA, a serem levadas em consideração para realizar uma partilha adequada dos dados de investigação.

O desenvolvimento até aqui foi quase sempre trilhado enfatizando os benefícios da CA, mas é preciso também estar atento as dificuldades que podem influenciar na partilha dos dados em AA. Portanto, outro fator que deve ser considerado são os impedimentos que

surtem como desafios para este tipo de partilha. Tais desafios, impedimentos e dificuldades que foram percebidos nas referências, mas não aprofundados, são tratados na próxima seção em termos de tensões e conflitos sob o acesso aberto, no ambiente da universidade.

3.4. TENSÕES E CONFLITOS SOB O ACESSO ABERTO

O AA, apesar de imprescindível para a CA e para a partilha dos dados, gera tensões, debates e enfrenta desafios que põem no centro da discussão o investigador, e este sob um novo paradigma da comunicação científica e das práticas de investigação – o acesso aberto. Neste novo paradigma, segundo Autran e Borges (2014), há uma ruptura do modelo tradicional de comunicação da ciência e a emergência deste novo modelo. Tal ruptura gera tensões e conflitos que fazem com que os investigadores lidem com inseguranças e barreiras de ordens diferentes em suas práticas de investigação. Como afirma Kuhn (1996) “Because it demands large-scale paradigm destruction and major shifts in the problems and techniques of normal science, the emergence of new theories is generally preceded by a period of pronounced professional insecurity” (p. 67).

Na literatura, os estudos sobre as atitudes, percepções e comportamentos dos investigadores em relação ao AA, podem fornecer uma visão geral sobre as tensões e conflitos²² que os envolvem. São estudos que em sua maioria, destacam as razões para o investigador realizar ou não a partilha e sobre as vantagens e as desvantagens que podem influenciá-los (Ferguson et al., 2014; Rowley et al., 2017). No mesmo sentido, outros trabalhos propõem descobrir mudanças nas atitudes dos investigadores ao longo do tempo (Gaines 2015; Xia 2010), ou encontrar diferenças em seus comportamentos conforme as áreas disciplinares (Lämmerhirt, 2016; Severin et al., 2018) e ainda, compreender as influências de atores externos, como os do sistema de comunicação científica, na conduta dos investigadores (Freitas, 2015; Freitas & Leite, 2019). Assim, trilha-se nesta seção tais perspectivas, de modo a construir uma visão geral sobre as tensões e conflitos que influenciam e devem ser alvos de consideração para realizar uma partilha adequada dos dados de investigação, no contexto do acesso aberto e sob o ambiente da universidade.

²² Segundo a infopédia.pt, a tensão representa o pródromo de um conflito, já o conflito está na perspectiva da coexistência de tendências, ou impulsos antagónicos.

Ao considerar o ambiente da universidade, o ponto de interesse aqui são as tensões que podem gerar conflitos e que têm potencial para interferir nas ações dos investigadores para uma tomada de decisão sobre como, e se deve ou não partilhar os dados em AA.

A dinâmica do investigador na universidade envolve práticas em um contexto multidisciplinar e de necessários relacionamentos com diversos atores, dentro e fora do ambiente da investigação e da universidade. Neste contexto, o próprio investigador é o ator central, isto porque cabe ao investigador, a priori, a responsabilidade pelos relacionamentos e a tomada da decisão da partilha e de como realizá-la. O problema, como já feito alusão, são os fatores que geram tensões e conflitos aos quais os investigadores estão sujeitos e que podem impedir ou interferir na partilha dos dados em acesso aberto.

Sob o contexto referido acima, o ambiente específico da investigação na universidade proporciona um intercâmbio sociotécnico, onde investigadores interagem entre si e com vários outros atores. Estas interações, somadas a necessidade de uso dos princípios e práticas da CA para a partilha em AA, formam um conjunto de fatores que potencializam as dificuldades para a disponibilização dos dados em AA. No entanto, este contexto não apresenta apenas dificuldades, mas também traz benefícios. Portanto, fatores positivos e negativos podem ser observados na relação de influência dos investigadores na universidade. Os fatores positivos foram em parte, citados nas seções anteriores, os negativos também, embora menos citados, estes últimos são os que tendem a gerar tensões que, contrastados com os fatores positivos podem resultar nos conflitos.

A partir do exposto, procura-se aqui construir uma visão geral sobre as tensões e conflitos que possam ser parte das considerações para realizar uma partilha adequada dos dados, permitindo desenvolver ações que gerem confiabilidade e segurança. Nesse intento, busca-se destacar o seguinte na literatura: a) os fatores influenciadores positivos e negativos que possibilitem a compreensão das tensões e conflitos; b) a cultura da partilha, nas áreas do conhecimento da investigação; e, c) a influência dos atores do sistema de comunicação científica na relação com os investigadores. Em todos os sentidos, procura-se conhecer de modo teórico as tensões e conflitos no que diz respeito as influências para a partilha dos dados em AA, no ambiente da universidade.

Em relação aos fatores influenciadores, positivos e negativos, Veiga (2017) ao investigar a percepção e o comportamento dos investigadores, seus estímulos e as barreiras para a partilha dos dados de investigação e de artigos científicos no AA, compila sob diversas referências²³, as variáveis e os fatores que podem influenciar positivamente ou dificultar a adesão dos investigadores na questão da partilha. Nesse sentido, os benefícios referenciados por Veiga, estão assim organizados e listados, seguidos da ideia principal:

- a) Acessibilidade – produção preservada e acessível em longo prazo;
- b) Visibilidade – contribuição para o aumento da taxa de citação;
- c) Credibilidade – maior disponibilidade para a validação por pares;
- d) Recompensa acadêmica – resultado de promoção pela maior visibilidade;
- e) Reconhecimento profissional – por pares e ocasionado pela partilha; e
- f) Altruísmo – motivação para a partilha em benefícios de outros.

Segue abaixo os fatores negativos, questionamentos ou receios que preocupam o investigador, referenciados por Veiga (2017), juntamente com a ideia principal e o motivo para tensão e/ou conflito que pode ser provocado pela partilha em AA:

- a) *Copyright* - infringir os acordos assinados com os editores científicos. A tensão encontra-se quando o material é partilhado primeiro em periódicos com acordos de *copyright*, há receio de que a partilha em AA desrespeite os acordos;
- b) Perda de oportunidade de publicação – antecipação de ganho e eliminação da vantagem competitiva. A tensão está no temor de que outros investigadores utilizem o material partilhado primeiro e publiquem ou comercializem com ganho antecipado;
- c) Informações sigilosas – partilha de dados sensíveis e reservados. O conflito é entre a necessidade da partilha e a confiabilidade da proteção da identidade

²³ Autores citados por Veiga (2017): Alle, 2005; Amante, 2013; Borges, 2006; Borgman, 2006, 2007,2009; Borrego, 2017; Campbell, 2002; Chan, 2004; Cragin, 2010; Crown, 2002; Davis; Connolly, 2007; Foster & Gibbons, 2005; Lane & Shur, 2010; Piowar, 2010 Polydoratou, 2008; Reidpath, 2001; Rodrigues, 2014; Savage & Vickers, 2009; Shearer, 2003; Swan & Brown, 2005; Van House, 2002; Van Raan, 1997.

e/ou a anonimização dos dados sensíveis e sigilosos e, nos termos e consentimentos, quanto à legalidade na reutilização dos dados;

- d) Tempo adicional e esforço – é a percepção de que atividade adicional diminui o tempo das atividades “principais” da investigação. É a ideia de que a atividade adicional reduz o tempo de outras, percebidas como mais importantes para a investigação, também é visto como uma sobrecarga negativa;
- e) Desconfiança/plágio - uso inadequado dos dados partilhados e a interpretação incorreta deles. É a tensão pela desconfiança quanto ao uso do que será feito da produção disponibilizada e a realização de cópias sem os devidos créditos;
- f) Usabilidade – falta de conhecimentos, facilidade ou formação para uso da tecnologia. A tensão está na dificuldade da aplicação das ferramentas tecnológicas, mas também na falta de facilidade para o uso pelo investigador;
- g) A cultura da partilha – resistência por ser incomum na área do conhecimento. O conflito está na tradição disciplinar em não partilhar, em questões geracionais e nas atitudes de pares que veem com desconfiança tal prática;
- h) Desconhecimento - não saber onde partilhar. O conflito está na intenção da partilha no ambiente do AA, mas falta o conhecimento sobre a existência ou possibilidades para a partilha no contexto trabalhado; e
- i) Influência das políticas mandatórias - pouca eficácia e garantias das políticas. O conflito está na obrigação da partilha e a tensão entre a resistência à obrigatoriedade e a não percepção das proteções para esta mesma partilha.

A autora ainda classifica o comportamento dos investigadores em três categorias: custos, benefícios e fatores contextuais, onde este último é subdividido em fatores individuais que remetem obviamente ao próprio indivíduo e, em fatores institucionais, que remetem a relação com as organizações. A respeito dos comportamentos individual e institucional, Kim e Stanton (2016) referenciando outra gama de autores²⁴, direcionam à

²⁴ Autores citados por Kim & Stanton (2016): McCullough, McGeary, & Harrison, 2008 ; Piwovar & Chapman, 2008; McCain, 1995; Louis et al., 2002; Bietz, Baumer, & Lee, 2010; Hey & Trefethen, 2004; Karasti, Baker, & Millerand, 2010; Choudhury, 2008; Witt, 2008; Campbell & Bendavid, 2003; Campbell et

possibilidade de uma categorização semelhante à de Veiga (2017), em três grupos: a) institucionais, como, por exemplo, as políticas das agências de financiamento, os requisitos de periódicos ou os contratos com os patrocinadores da indústria; b) técnicos, como os recursos de tecnologia da informação e os repositórios de dados; e, c) individuais, constituído pelas características pessoais, benefício, esforço e os riscos percebidos pelos participantes.

Geralmente os estudos sobre a temática convergem para os fatores e categorias listados acima, atribuindo designações diferentes e algumas vezes destacando aspectos mais gerais ou enfatizando determinados detalhes. Para uma visão geral, a tabela 11 mostra que as categorias que classificam o comportamento contém fatores negativos geradores de tensões e, que as tensões, quando contrastadas por fatores positivos são potenciais geradoras de conflitos. Ambos os fatores podem influenciar na partilha dos dados em AA.

Tabela 11

Categorias de comportamentos geradoras de tensões e conflitos sob o acesso aberto

Categorias	Exemplo de fatores negativos geradores de tensão	Exemplos de fatores positivos ou necessidades que podem gerar conflito a partir de uma tensão
Individual	Perda de oportunidade	- Ganho de oportunidade de visibilidade e citação;
	Esforço adicional	- Recompensa acadêmica resultante da maior visibilidade;
	Desconfiança/plágio	- Reconhecimento pelos pares e credibilidade pela partilha;
	Cultura de não partilhar	- Adaptação e atualização nas transformações da ciência;
	Desconhecimento	- Desejo altruísta para benefícios sociais e científicos;
Institucional	Copyright	- Acesso legal e ético com claros códigos de conduta;
	Políticas mandatárias	- Interesse em financiamentos e investimentos;
Técnico	Informações sigilosas	- Segurança e rapidez no tratamento dos dados; e
	Falta de usabilidade	- Acesso seguro, rápido e eficiente.

Portanto, faz-se necessário observar o quadro acima para entender que os fatores negativos e positivos podem estar relacionados entre si gerando tensões e conflitos. Estes, por sua vez, criam dificuldades cognitivas e comportamentais para a realização da partilha em AA: por exemplo, a intenção do investigador em não partilhar, gerado por uma expectativa de perda de oportunidade, pode resultar em um conflito justificado pelo desejo da partilha para obter uma maior visibilidade. Em outro sentido, o maior número de citações

al., 2002; Kankanhalli, Tan, & Wei, 2005; Kim, 2007; Kling & Spector, 2003; Tenopir et al., 2011; Reidpath & Allotey, 2001; Savage & Vickers, 2009.

para o investigador pode não ser motivo suficiente para partilhar, quando contrastado com o esforço adicional despendido. Ou ainda, a perda de tempo talvez não seja relevante, quando ponderado com os benefícios do armazenamento e promoções pessoais. Isto significa que nem fatores negativos, nem positivos podem ser vistos isoladamente, pois estão intimamente relacionados ao indivíduo e ao contexto.

Em relação ao contexto, a falta de infraestrutura que demande um esforço adicional dos investigadores, ou a falta de esclarecimentos institucionais, ou ainda uma política de recompensas inexistente, são fatores ambientais que podem influir na partilha. Assim, para partilhar de forma mais segura e confiável em AA, as especificidades do ambiente também devem ser consideradas. Sobre o fator ambiente, que, no caso deste trabalho, é a universidade, a literatura consultada destaca mais duas potenciais fontes geradoras de tensão e conflitos a serem consideradas: as diferenças na cultura da partilha por áreas disciplinares e a participação dos atores do sistema de comunicação da investigação.

As diferenças na cultura da partilha no comportamento dos investigadores de acordo com a área de conhecimento, ou com as práticas de investigação, são moldadas nas tradições (Harley et al., 2010). Por exemplo, em disciplinas que lidam com os seres humanos, as preocupações com a privacidade dos pacientes ou entrevistados, bem como as regulamentações legais, podem ser barreiras adicionais à partilha (Institute of Medicine [IOM], 2013), em áreas disciplinares das ciências exatas e naturais, quando comparada a outras áreas, tradicionalmente disseminam os resultados das investigações em artigos de periódicos, incluindo a publicação em AA (Costa & Leite, 2017). Portanto, a receptividade e a atividade disciplinar para a partilha pode variar de acordo com a cultura em áreas específicas do conhecimento (Antelman, 2004, 2006; Swan & Brown, 2005; Thessen & Patterson, 2011).

Knorr-Cetina e Reichmann (2015), usam a noção de culturas epistêmicas para explicar e contrastar as diferenças de domínio nos processos de produção de conhecimento, trazendo à tona não apenas o nível da prática, mas um nexos que se desenvolvem dentro de uma especialidade e ambientes ricos e complexos, que de acordo com os autores, tendem a impor e expandir as suas próprias estruturas e preocupações, mas que também podem ser dilacerados por conflitos e desentendimentos e pela convivência nada pacífica dos participantes

(humanos e não humanos). Os autores ainda afirmam que é preciso entender e observar “o sistema” e como ele possibilita e restringe a ciência, alertando ser mais provável investigar:

[...] how practitioners distinguish signal from noise, or move from in vitro experimentation to in silico simulation, or how they decide which figure to trust when experimental outcomes are uncertain, than on these practitioners’ budgets, career moves, or conflicts with management (p.873).

Em uma abrangente investigação sobre as práticas e percepções de cientistas em relação a partilha, a gestão e a reutilização de dados, conduzida entre várias disciplinas diferentes com mais de dois mil investigadores de oitenta e sete países, Tenopir et al. (2020) obtiveram conclusões e verificaram diferenças significativas entre áreas específicas do conhecimento, para o seguinte: a) na disposição em partilhar os dados; b) no uso dos dados coletados por terceiros; c) nas práticas de armazenamento de dados; d) no que diz respeito ao trabalho com bibliotecários e gestores; e) em relação ao treinamento e assistência para a RDM, fornecida pela organização; e, f) na satisfação com as ferramentas disponíveis.

Ao considerar a cultura epistêmica, no sentido de que a partilha por áreas específicas do conhecimento é indissociável da investigação no ambiente da universidade e que esta cultura tem influências positivas e negativas nas atitudes dos investigadores, pode-se perceber que, além da tradição, também as particularidades e a abrangência dos temas que as investigações nas áreas do conhecimento proporcionam, podem resultar em novas tensões e conflitos que só se aplicam àquela disciplina. Assim, cabe atentar para as especificidades das áreas do conhecimento como fator essencial para minimizar as tensões e conflitos oriundo das atividades para a partilha dos dados em AA na universidade.

Outro potencial fator gerador de tensões e conflitos a ser considerado, é o sistema de comunicação científica que interage com o investigador e que envolve outros atores no contexto da investigação na universidade. Conforme coletânea adotada por Freitas (2015)²⁵, podem ser considerados como atores que constituem tal sistema: o investigador, as editoras, as bibliotecas, as agências de fomento e os gestores das universidades. A autora considera que estes atores são os mais influentes no contexto da comunicação da investigação e que a

²⁵ Autores citados por Freitas (2015): Antelman (2004); Björk (2007); Bethesda Statement on Open Access Publishing (2003); Björk e Solomon (2014); Meadows (1999); Ohira (1998); Pinfield (2010).

relação entre tais atores no contexto da comunicação científica, é fonte para a verificação de tensões e conflitos que podem interferir na partilha em AA. Nesse sentido é essencial que a produção científica esteja relacionada com a avaliação, acesso, visibilidade, preservação e a interoperabilidade (Martins, Nolasco & Silva, 2013).

Freitas e Leite (2019) em estudo realizado sobre percepções em relação ao depósito em repositórios de AA, perguntaram aos investigadores se poderiam perceber uma possível influência, positiva ou negativa, por parte de atores do sistema de comunicação científica sobre as decisões da partilha. Os resultados constataram que quase todos os investigadores citaram influências negativas. Ainda segundo os mesmos autores (Freitas e Leite, 2019a), os principais atores do sistema de comunicação científica e as influências em relação a partilha, são as seguintes:

a) Editoras: a influência encontra-se nos direitos cedidos a estas instituições, que determinam o que podem fazer com a obra em termos de partilha. Algumas editoras, na tentativa de se adequarem à tendência do AA ou para implementar um novo modelo de transferência de custos, passam a cobrar uma taxa de processamento dos artigos (*Article Processing Charge - APC*)²⁶, assim o investigador fica livre para disponibilizá-lo em AA. O conflito encontra-se no entendimento de que esse modelo resulta em uma má qualidade das publicações em AA, pois devido ao pagamento da taxa, critérios qualitativos seriam ignorados resultando em resistência para a partilha²⁷ (Beall, 2012; Björk & Solomon, 2014).

²⁶ As taxas de processamento de artigos (APCs) constituem um modelo de viabilidade financeira de novas editoras em acesso aberto, como a Public Library of Science, que cobra aos autores e não aos leitores, como acontece no modelo convencional, os custos associados à publicação. A cobrança de APCs foi aproveitada também pelas chamadas “editoras híbridas” que combinam no mesmo fascículo artigos em acesso aberto e artigos fechados, conforme o pagamento, ou não, das APCs dos artigos. No entanto, nem todas as publicações em acesso aberto cobram APCs, como pode ser concluído dos títulos em acesso aberto que constam do Directory of Open Access Journals (DOAJ): a 1 de dezembro de 2021, dos 17190 títulos do DOAJ, 12151 (71%), não cobravam APCs (<https://doaj.org/>). O maior problema prende-se com as editoras predadoras que cobram APCs, mas sem conferir ao processo editorial, particularmente no que concerne à arbitragem científica, o rigor que deveria ter, causando danos à imagem de periódicos de acesso aberto que cobram APC (Nassi-Caló, 2015).

²⁷ Outra possibilidade já citada anteriormente, relaciona-se com o *Copyright* e o receio em infringir os acordos assinados com os editores científicos quanto aos direitos do autor da publicação em acesso aberto (Antelman, 2006; Veiga, 2017).

b) Bibliotecários: o estudo aponta para a falta de diálogo e o desconhecimento do papel destes profissionais. Nesse sentido, foi proposta a criação de funções para desenvolvimento de mecanismos de apoio aos investigadores na transição das publicações para o AA, com a divulgação destes mecanismos para a comunidade científica. Segundo Cooke (2007), por questões de tempo e esforço, os investigadores preferem uma publicação mediada em AA, porém a indicação é que as bibliotecas e bibliotecários não estão, em sua maioria, preparados em estrutura e capacitação para um apoio amplo, de esclarecimentos e mediação a nível da universidade (Antelman, 2004; Joint, 2006).

c) Agências de fomento: a tensão aqui é direcionada para o fato das agências exigirem que as publicações e os dados estejam disponíveis para a partilha. Os conflitos e as tensões gerados pelas agências de fomento talvez sejam um dos fatores mais citados nos estudos, tendo em vista as políticas de AA para investigações financiadas com dinheiro público e, em contrapartida, a exigência de que os investigadores disponibilizem suas produções em AA, isto envolve barreiras por parte dos investigadores com influências positivas e negativas para a partilha.

d) Universidades: os investigadores respondentes do estudo, declaram que há falta de esclarecimento e de políticas por parte dos gestores a nível de reitoria, direção ou coordenação dos cursos. Segundo os autores, os gestores da universidade são responsáveis por garantir a infraestrutura científica e tecnológica para realizar, gerir e promover as investigações, bem como estabelecerem as políticas institucionais, inclusive para a partilha em AA. Portanto, além da influência da universidade como gene da produção do conhecimento, a depender do conhecimento dos gestores e do envolvimento da universidade com o AA em termos de ações, como por exemplo, o sistema de recompensas, o nível de tensão e conflito pode ser maximizado ou minimizado, tendo em vista a influência que os atos da gestão e dos gestores podem exercer na investigação e nos investigadores.

e) Investigadores: como produtores e geradores de conhecimento no ciclo da comunicação científica e no processo de investigação, ao mesmo tempo consomem e comunicam tal conhecimento. Por serem parte do sistema são influenciados pelos demais atores, mas também exercem um papel primordial de influência, pois possuem o poder de decidir sobre a forma de disponibilização de seus trabalhos em acesso aberto.

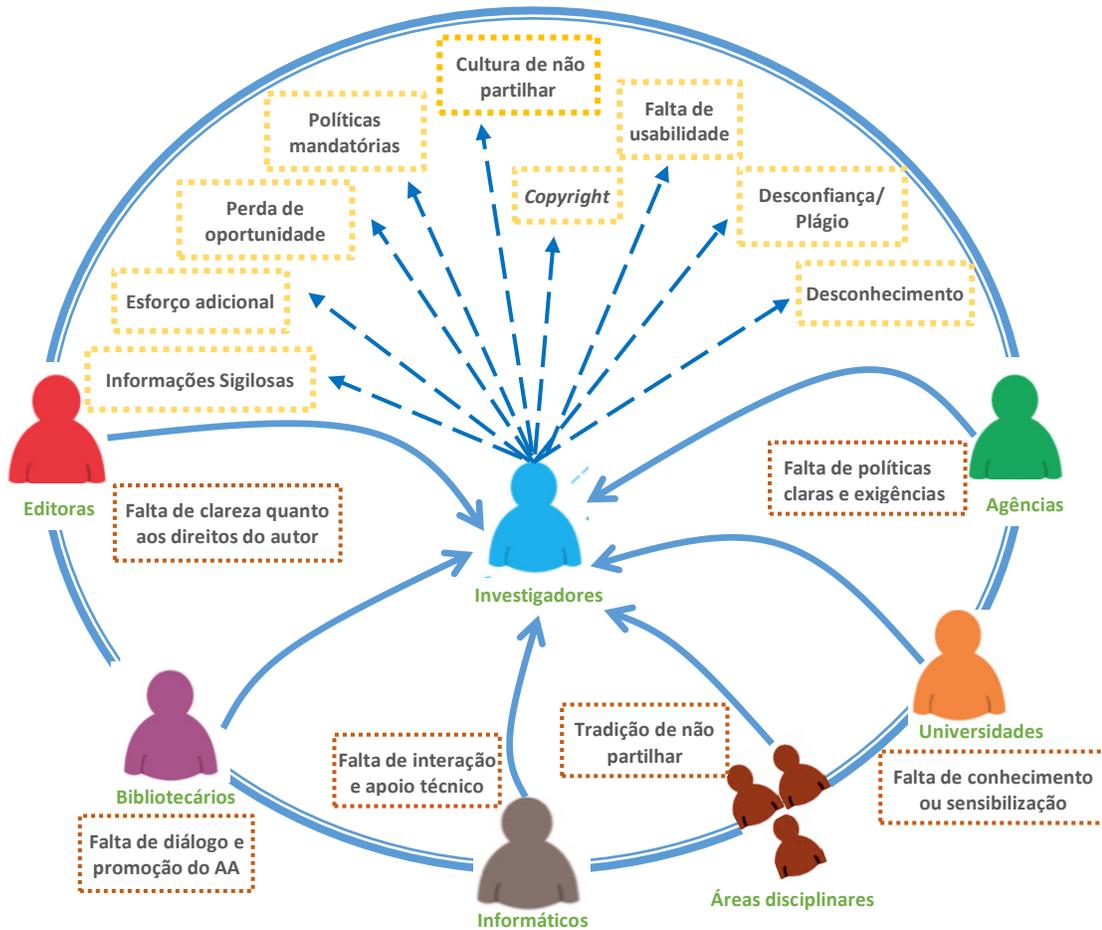
É possível pensar nos atores do sistema de comunicação e na cultura da partilha das áreas disciplinares como influências ambientais externas, que podem repercutir de modo interno e individual no investigador. Isto porque, ao longo dos anos, a literatura tem mostrado consistentemente que a maioria dos investigadores é a favor da partilha em AA (Dallmeier-Tiessen et al., 2011; Tenopir et al., 2017; Wallis, Rolando & Borgman, 2007) e que, embora haja tensões e conflitos, os investigadores estão claramente interessados e partilhar em AA (Piwowar et al., 2018). Logo, espera-se que a partilha se manifeste de forma rotineira. No entanto, apesar da postura positiva, a prática mostra que isso nem sempre acontece na intensidade e na proporção do interesse que é demonstrado, pois os investigadores parecem relutantes em adotar a partilha em AA (Rowley et al., 2017).

Borgman (2012) afirma que a maioria dos investigadores respondem positivamente e previsivelmente ao “querer partilhar” como uma inserção pró-social em face do movimento crescente sobre o tema e o que é considerado bom para a comunidade científica, nem sempre é percebido de modo igual para o investigador. As afirmações ratificam a ideia que faz da prática da partilha um dilema, um conflito, pois as atitudes e comportamentos dos investigadores, nesse caso, não correspondem ao *ethos* coletivo e geral da ciência, mas podem ser guiados por interesses individuais (Scheliga & Friesike, 2014). Neste sentido, os investigadores também podem agir com base na solidão, particularismo (busca de crédito), interesse (próprio) e dogmatismo (fé nas próprias descobertas). Mitroff (1974) e Mulkey (1976), já destacavam a necessidade da reflexão sobre a ambivalência entre o "ideal" dos investigadores e o comportamento real como parte do empreendimento científico.

Nesta seção, possíveis fontes de tensões e os conflitos foram observados a partir de três aspectos: 1) as preocupações dos investigadores; 2) as diferenças na cultura da partilha; e, 3) as influências de atores externos sobre o investigador e a investigação. Percebe-se que os fatores de tensão que geram conflitos podem ser interrelacionados, reforçando a importância do contexto ambiental e do olhar articulado para a partilha dos dados em AA na universidade. A figura abaixo propõe a observação das influências em relação ao investigador, ou seja, são questões relacionais, provenientes de inquietações da individualidade e questões da coletividade que repercutem de dentro para fora e de fora para dentro, podendo afetar a atitude do investigador para partilhar os dados em acesso aberto.

Figura 6

Fatores que podem repercutir na decisão do investigador na partilha em acesso aberto



Nota. Preocupações e influências capazes de gerar tensões e conflitos (adaptado de Freitas e Leite, 2019).

Portanto, é importante para a universidade ou para qualquer tipo de planejamento que aborde a partilha dos dados em AA, levar em consideração as questões cognitivas, juntamente com as questões sociais, técnicas e institucionais para partilhar os dados de investigação de forma adequada. Em uma abordagem relativamente nova e com diversos fatores ainda por esclarecer, a adoção da partilha dos dados em AA exige uma mudança no comportamento dos investigadores no que diz respeito à forma de realizar as investigações e no processo que o envolve (Pontika et al., 2015). Com isso, deve-se considerar os problemas decorrentes das tensões e conflitos e como eles se relacionam, tanto para permitir esclarecimentos e direcionamentos consistentes, quanto para que os investigadores tenham confiabilidade e segurança para partilhar adequadamente os dados de investigação em AA.

3.5. A ABERTURA DOS DADOS DE INVESTIGAÇÃO

Até agora a partilha em AA foi abordada de modo geral, principalmente em termos de disponibilização da produção científica, sem distinções intencionais entre os dados e os outros materiais digitais relevantes para a investigação. Porém, são os dados o objeto da partilha da RDM na universidade. Assim, é necessário observar questões concernentes aos dados abertos, objetivado em conhecer e formar uma visão geral para a abertura dos dados de investigação e conseqüentemente, constituir o que deve ser considerado para realizar uma partilha adequada dos dados de investigação, no ambiente da universidade.

Como já referido, os dados de investigação para a partilha em AA são objetos informacionais distintos dos demais elementos da produção científica, por sua complexidade, diversidade e pela variabilidade que podem ter os seus significados. E que, a relação entre a CA, a abertura e o AA estão conceituadas de forma interdependente (Willinsky, 2005). Assim, respeitando as particularidades de cada elemento, percebe-se a possibilidade de tal convergência através dos “dados abertos”²⁸.

Os dados abertos podem ser entendidos a partir da seguinte definição: “is the practice of making underlying research data publicly available, accessible and reusable with minimal restrictions. Within the broader shift towards open Science” (Higman, Bangert & Jones, 2019). Ratifica-se que na partilha pública dos dados em AA, a “abertura” é o processo para o alcance deste tipo de partilha e, que os dados abertos juntamente com o AA são pilares da ciência aberta, portanto interrelacionados para a abertura dados de investigação.

A “Declaration on access to research data from public funding”, idealizada por países membros da OCDE em 2004, divulgada em 2006 e atualizada em 2021 (OCDE, 2021), juntamente com o relatório da Royal Society (2012), são exemplos de importantes documentos orientadores sobre o tema da abertura e dos dados abertos de investigação. Tais documentos abordam a necessidade de adaptação às mudanças proporcionados pelos cenários tecnológicos, social e político, reunindo recomendações aos investigadores, universidades, institutos de investigação, agências de fomento, governos, editoras de

²⁸ Neste estudo o termo “dados abertos” refere-se aos dados de investigação, pois este termo pode ser referido de modo geral ao contexto governamental; porém, o conceito e sua utilização envolvem um espectro amplo de categorias e aplicações no campo da ciência aberta (Albagli, Clinio, & Raychtock, 2014).

revistas científicas e organismos profissionais (Jones, Pryor & Whyte, 2013). Mas também, a partir da BOAI em 2002 e da declaração de Berlim em 2003, os dados de investigação e os repositórios evoluem como importantes recursos abertos, que juntos, são direcionados por políticas, práticas para a abertura, partilha e reutilização dos dados (Pampel & Dallmeier-Tiessen, 2014).

Neste contexto que envolve mudanças, políticas, orientações e vários atores para a abertura dos dados, dois itens podem ser considerados para formar uma visão geral da abertura dos dados, objetivadas nesta seção. O primeiro, são as infraestruturas tecnológicas que viabilizam principalmente a gestão e a partilha dos dados, representado pelos repositórios. O segundo, é a possibilidade da reutilização dos dados. Isto porque ambos estão dentre os propósitos centrais da RDM que permitem a abertura dos dados, mas também no resultado esperado para depois da partilha dos dados de investigação.

Os repositórios digitais, como infraestruturas tecnológicas²⁹, podem ser considerados a melhor opção para garantir a gestão, a visibilidade e a disponibilidade dos dados para um maior público (Joint Information Systems Committee [Jisc], 2011). Entre variações terminológicas, como “repositórios científicos”, “centros de dados”, ou “repositórios de dados” (Maron & Smith, 2008), estas plataformas caracterizam-se de acordo com o tipo de objetos que albergam, podendo, assim, serem institucionais, disciplinares ou temáticos, multidisciplinares e orientados por projetos (Pampel et al., 2013). Uma definição mais comum, considera o objetivo da construção do repositório. Ou seja, os classifica em repositórios institucionais – voltados para a o armazenamento da produção científica de uma determinada instituição (Lynch, 2003) e repositórios temáticos, voltados para a reunião da produção técnico-científica em determinada temática (Kuramoto, 2006). Cabe ressaltar que outras nomenclaturas ou infraestruturas digitais podem ser utilizados sob o mesmo propósito, tais como projetos ou repositórios colaborativos disponíveis na *Web*, a exemplo do *GitHub*.

²⁹ São exemplos de repositórios: GenBank, Dryad, Figshare e Zenodo, bem como o Dataverse e o Mendeley Data. Repositórios de dados temáticos: Knowledge Network for Biocomplexity (KNB) e o repositório protocols.io, o GitHub - usado para projetos que envolvem programação, softwares e código-fonte aberto e o Institute of Electrical and Electronics Engineers (IEEE). A maioria desses repositórios oferecem dados e investigações abertas disponíveis para acesso e reutilização (Piwowar et al, 2008).

A partilha através dos repositórios digitais é aparentemente a forma mais usual para a realização da abertura dos dados, podendo esta partilha ser realizada como parte integrante dos artigos e periódicos, ou como arquivos suplementares (Reilly et al., 2011). Assim, muitos periódicos em AA solicitam como condição de publicação que os autores depositem os dados em repositórios digitais públicos, para serem acompanhados e vinculados ao artigo submetido (Bechhofer et al., 2013). Publicações em “data papers” - artigos que descrevem um ou vários conjuntos de dados acessíveis *online* - e, “data journals” - periódicos, cuja finalidade é descrever os dados de uma investigação -, geralmente exigem o depósito em repositórios de dados abertos (Chavan & Penev, 2011). As políticas e mandatos desenvolvidos por fontes institucionais ou financiadas pelos governos, como a *National Science Foundation* (NSF, 2011) e o *National Institutes of Health* (NIH, 2003), também requerem dos beneficiários a documentação e a partilha em repositórios de dados.

Segundo Kowalczyk e Shankar (2011), o repositório de dados é definido como um sistema que provê um conjunto de serviços para o acesso e, ao mesmo tempo, para o arquivamento dos dados. Os autores declaram que se deve obedecer a três aspectos: contexto, fixidez e persistência³⁰. Assim, é necessário que estes repositórios possam guardar mais do que dados, mas permitam armazenar “workflows”, modelos, códigos e cadernos de laboratório. Ou seja, um conjunto de objetos relevantes para a posterior reutilização.

Alguns benefícios devem caracterizar os repositórios digitais. Sayão e Sales (2016) listam os mais perceptíveis:

- Partilha - agrega e organiza os recursos informacionais no tempo e no espaço;
- Visibilidade, permite que os dados sejam consultados e citados;
- Preservação digital - guarda em longo prazo os dados de valor contínuo;
- Créditos ao autor - permite que os autores sejam reconhecidos, citados, avaliados e recompensados pelo trabalho;

³⁰ O primeiro aspecto refere-se à capacidade do repositório de fornecer contexto suficiente e informações sobre a proveniência e a descrição dos dados para garantir a descoberta e a fácil assimilação do conteúdo dos dados. A fixidez corresponde à capacidade do repositório em verificar se os arquivos digitais mantidos estão inalterados e não corrompidos. A persistência refere-se à capacidade do repositório de dados garantir o arquivamento em longo prazo e o acesso contínuo aos dados.

- Segurança - armazena os dados de modo confiável com garantias físicas e lógicas;
- Disponibilidade - garante que os dados estejam on-line para serem acessados por pessoas ou por sistemas;
- Curadoria - possibilita os processos promovidos pela curadoria digital;
- Redes de repositórios – maximiza o trabalho por meio da interoperabilidade; e
- Reutilização - aumenta o grau de reinterpretação dos dados, possibilitando novas investigações interdisciplinares.

Como pode-se perceber, os repositórios permitem pôr em prática vários princípios da CA. Portanto, os princípios e práticas da CA são base para uso desta ferramenta. No entanto, para resultar no avanço e nos benefícios da abertura, sob a CA, os investigadores devem partilhar os dados de modo a garantir a sua reutilização. Assim, sob os princípios e práticas da CA, os repositórios têm sido desenvolvidos para armazenar, organizar, preservar e oferecer acesso aos dados para posterior reutilização (Marcial & Hemminger, 2010).

A disponibilização dos dados em repositórios digitais de AA não garante os benefícios da abertura, antes pode resultar nos investigadores uma sensação de “cumprimento de dever”, seja pelas pressões institucionais via mandatos, ou motivado por aspectos positivos e iniciativas estabelecidas na comunidade científica, o próprio termo “depósito” ou “arquivamento” aqui já utilizado, pode dar um falso sentido finalístico ao processo³¹. Soma-se a isso que o contexto atual da investigação para a partilha de dados em AA tende a tornar-se uma prática sistemática, ou seja, conforme mais instituições adotam e/ou implementem políticas e incentivos sobre o AA, mais investigadores são incitados a abrirem os seus dados, tornando-os amplamente disponíveis em formato digital. Por essa tendência, os incentivos e as preocupações frequentes com a reutilização dos dados partilhados necessitam ser observadas para não incorrer no risco do desuso, com dados ociosos em repositórios e permitindo criar indesejados “cemitérios de dados”.

³¹ Em contrapartida ao termo “partilha”, entende-se que o “depósito” ou “arquivo” pode ser percebido apenas no sentido da guarda do material, o que não expressa toda a dinâmica necessária para uso posterior e necessário dos dados de investigação, principalmente no contexto objetivado para a reutilização destes dados.

Em outro sentido, a prática da partilha dos dados em repositórios digitais, quando adequadamente partilhados e posteriormente reutilizados, não apenas possibilita pôr em prática os benefícios da CA, mas também permite transformar os resultados de estudos anteriores em novas investigações (Fienberg, Martin & Straf, 1985; Zimmerman, 2008). É principalmente sob a viabilidade do uso dos dados para as novas investigações, que está bem fundamentada a ideia da reutilização³². Assim, o processo para o reaproveitamento na ciência pode ser o resultado da intensificação das iniciativas da CA e dos dados abertos, expandidos com a possibilidade da partilha e na utilidade dos repositórios (Marcial & Hemminger, 2010). Portanto, o objetivo da reutilização, segue no sentido da descoberta dos dados para utilização atual e futura (Abbott, 2008), de modo a ser possível sua análise e uso de formas diferentes, com valor para investigadores atuais e vindouros. Segundo Heidorn, (2008), o verdadeiro valor dos dados não é determinado pelo custo da recolha, mas pela economia incorrida por alguém que não precisa reuni-los novamente.

No sentido exposto, a importância da reutilização dos dados centra-se no princípio de que os dados partilhados pelas investigações atuais também possam evoluir em usos futuros, como o exemplo de Dimitri Mendeleiev, idealizador da tabela periódica, que a organizou graças à reutilização de materiais de outros cientistas, disponíveis naquela época (Gordin, 2012). De forma semelhante, o tratamento para a gripe espanhola torna-se viável quando um tecido infectado pelo vírus foi encontrado em um hospital da Inglaterra, o que originou as investigações desta amostra, sendo possíveis graças à preservação e a partilha dos arquivos científicos de 1916, oriundos daquele hospital militar (Ditadi, 2003).

Em relação à abertura dos dados, Henning et al. (2019), entendem que “é possível afirmar que quanto mais abertos estiverem, mais serão usados, reutilizados e combinados com outros dados, promovendo o crescimento econômico, a inovação e o desenvolvimento” (p. 394). Ao referenciar Giglia e Swan (2012), os autores relacionam as vantagens em se ter os dados abertos. Nesse sentido, a tabela 12 mostra de forma sumarizada, o valor da abertura dos dados para a ciência e para a sociedade e onde seria possível a sua reutilização.

³² Sobre o conceito de reutilização veja, por exemplo, Pasquetto, I., Randles, B., & Borgman, C. (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16(0). <https://doi.org/10/gf87pf>

Tabela 12*Vantagens dos dados abertos e da reutilização para a ciência e para a sociedade*

Vantagens dos dados abertos e da reutilização para a ciência	Vantagens dos dados abertos e da reutilização para a sociedade
Acessar os dados para compará-los, contrastá-los e debatê-los. Pode-se ter uma ciência mais sólida baseada em dados.	Reutilização para inovação e aumento da competitividade. A abertura dos dados pode ajudar a gerar novos produtos, serviços e empregos.
Adicionar novos dados para criar novo conhecimento. Os dados abertos são mais visíveis, se propagam mais rapidamente, aceleram a criação do conhecimento e promovem o progresso da ciência.	Reutilização para a criação de valores por intermédio da mineração de dados. A exploração do vasto recurso de dados e informação pode gerar benefícios para a economia, aprimorando a produtividade e o valor adicional para o consumidor.
Reutilizar os dados para novas investigações sem custos para coletá-los novamente. Uma vez coletados, analisados e depositados, poderão ser usados em novas e diferentes perspectivas.	Reutilização para as novas mixagens. A abertura dos dados pode ser utilizada tanto em aplicações para computador e dispositivos móveis.
Reutilizar os dados para verificar e detectar a fraude e/ou a falsificação. Mais transparência poderá evitar fraude.	Reutilização para preservação. A preservação será mais fácil se os dados estiverem abertos e interoperáveis via padrões comuns.
Criar novos dados com técnicas da mineração de dados. Aplicável aos dados abertos e licenciados para reutilização, proporciona aumento de valor e economia de tempo.	Reutilização para formulação de tomada de decisão. Os formuladores de políticas podem ser capazes de tomar decisões baseadas em evidências e monitorar o impacto dessas decisões.
	Reutilização para o conhecimento. A abertura dos dados permite que os cidadãos sejam mais informados nos níveis político, cultural e científico.

Nota. Adaptado de Henning et al. (2019).

Apesar do potencial para a reutilização na ciência e na sociedade ser um parâmetro para a avaliação dos dados de investigação e um dos objetivos para a sua abertura, a reutilização em outros contextos é um problema complexo. Borgman (2007), afirma que tornar o conteúdo que foi criado para uma audiência útil para outra é desafiador. Isto porque cada disciplina tem seu próprio vocabulário, estrutura de dados, práticas de investigação, elaborando questões de formas distintas e usando as suas próprias terminologias.

Portanto, reutilizar os dados para aplicação em outros contextos, também possibilita a ocorrência de erros e geração de desvios semânticos substanciais (Piggot, Hobbs & Gammack, 2001), sendo mais facilmente observáveis nos casos onde a documentação de apoio esteja em falta ou seja insuficiente (Niu, 2009), pois esta fornece o significado, a estrutura e a identificação para que os dados possam ser reutilizados de modo adequado. Assim, disponibilizar os dados em repositórios sem a contextualização estrutural

e semântica, impossibilita a interpretação e a sua reutilização, bem como a transmissão do conhecimento em outros contextos e, conseqüentemente, reduz o valor para a investigação interdisciplinar e futura. Ou seja, os dados de investigação não têm valor sem os seus metadados³³ e documentações apropriadas que descrevam seus contextos, bem como as ferramentas utilizadas para criá-los, armazená-los, adaptá-los e analisá-los (Kindling, 2013). Desta forma, os dados precisam de informação representativa para se tornarem objetos informacionais (CCSDS, 2012) abertos, disponíveis e acessíveis em repositórios digitais que possibilitem a sua reutilização.

A reutilização também requer que os investigadores tenham a capacidade de descobrir e acessar dados inteligíveis, confiáveis e relevantes (Thessen & Patterson, 2011). Portanto, a abertura dos dados de investigação é centralizada na ideia de que, quando partilhados, estes dados precisam ser acessados e terem seus significados e estruturas recompostas. Para isso, os dados nos repositórios precisam ser encontrados, ou seja, os dados e suas versões hospedadas precisam ser capazes de serem localizados de forma contínua, o que é essencial para o acesso e para a reutilização. A localização pode ser viabilizada pelos processos de interoperabilidade e de interação com outros recursos, como, por exemplo, identificadores persistentes, controle de versões e citações padronizadas. Nesse sentido, em resposta às necessidades de localização, acesso, interoperabilidade e reutilização dos dados, os princípios FAIR são citados de modo recorrente na literatura.

Os princípios FAIR, instituídos em 2016 pelo grupo FORCE 11³⁴, visam implementar um conjunto de orientações definidas tanto para uso por mecanismos computacionais automatizados, quanto por pessoas. Estes, quando devidamente adotados viabilizam a interoperabilidade entre diferentes ambientes de dados. Os quatro princípios fundamentais são descritos de forma mais explícita e mensurável por quinze guias elaboradas para o fornecimento de orientações (Wilkinson et al., 2016), conforme a tabela 13. Embora os princípios FAIR possam ser observados e utilizados de maneiras diversas,

³³ Metadado é a informação estruturada que descreve, explica, localiza, ou possibilita que um recurso informacional seja fácil de recuperar, usar ou gerenciar (NISO, 2004).

³⁴ <https://www.force11.org/fairprinciples>

em uma perspectiva reduzida e centrada na reutilização dos dados, os três primeiros princípios podem ser vistas como colaboradoras para a reutilização objetivada.

Tabela 13

Princípios FAIR e guias orientadoras para os dados de investigação

Princípios FAIR	Guias orientadoras
Data should be Findable	F1. (meta)data are assigned a globally unique and persistent identifier (DOI) F2. data are described with rich metadata F3. metadata clearly and explicitly include the identifier of the data it describes F4. (meta)data are registered or indexed in a searchable resource
Data should be Accessible	A1. (meta)data are retrievable by their identifier using a standardized communications protocol A1.1 the protocol is open, free, and universally implementable A1.2 the protocol allows for an authentication and authorization procedure, Where necessary A2. metadata are accessible, even when the data are no longer available
Data should be Interoperable	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles I3. (meta)data include qualified references to other (meta)data
Data should be Reusable	R1. meta(data) are richly described with a plurality of accurate and relevant attributes R1.1. (meta)data are released with a clear and accessible data usage license R1.2. (meta)data are associated with detailed provenance R1.3. (meta)data meet domain-relevant community standards

Nota. Extraído de Medyckyj-Scott et al (2016).

Diversas iniciativas são realizadas para que os princípios FAIR sejam implementados. A OCDE (2019) reconhece o surgimento de um consenso sobre a necessidade de melhorar o acesso aos dados de investigação e torná-los FAIR. A Comissão Europeia lançou em 2016 o relatório “Realising the European Open Science Cloud” (Comissão Europeia, 2016b), da *Commission High Level Expert Group on the European Open Science Cloud* (EOSC), com o objetivo de acelerar a transição mundial para uma ciência inovadora, de modo a construir iniciativas em direção a interoperabilidade de dados e serviços com todas as instituições que seguirem os princípios FAIR.

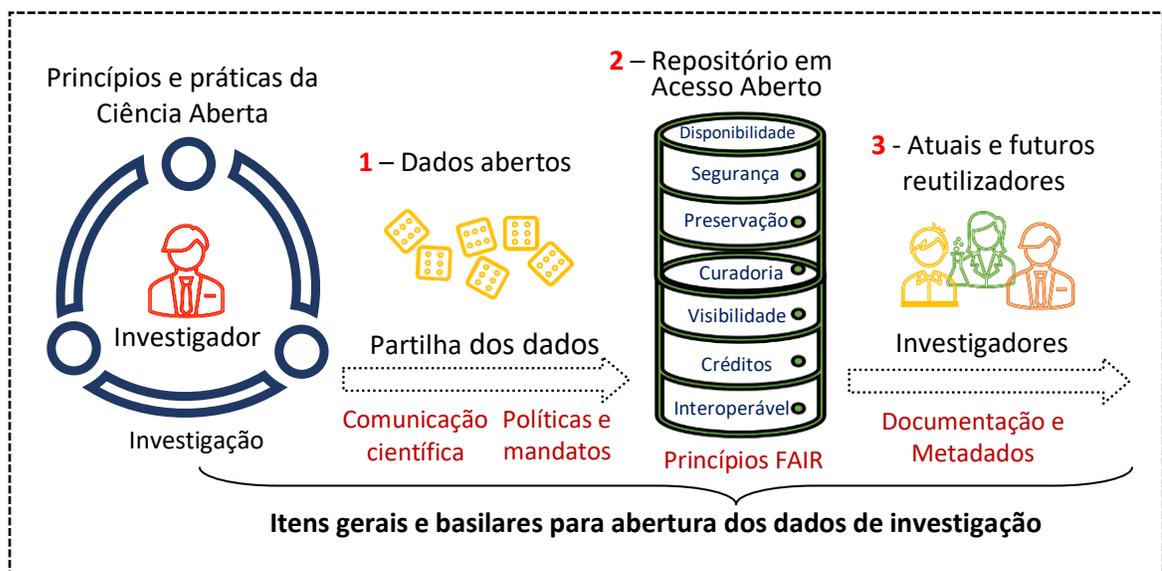
Embora os princípios FAIR venham sendo adotados de forma rápida por cientistas, editores, agências de financiamento e formuladores de políticas (Budroni et al., 2019), por

serem recentes no mundo científico, quando colocados em prática, podem gerar equívocos devido às dificuldades no entendimento e a falta de experiências na sua aplicação. Nessa direção, Dunning et al. (2017), avaliaram o alinhamento aos princípios FAIR em 38 repositórios digitais holandeses. Os autores relatam que todos estavam com baixa compatibilidade aos princípios FAIR. Assim, embora sejam adequados na reutilização dos dados, é pertinente considerar e avaliar requisitos, propósitos e necessidades para sua implementação, ou seja, é necessário um planeamento detalhado para a sua aplicação.

Nesta seção foi apresentada que a abertura dos dados significa torná-los disponíveis publicamente, acessíveis e reutilizáveis com restrições mínimas, direcionados pela CA e que, para abrir os dados tendo por base os princípios e práticas da CA, duas práticas são consideradas fundamentais: a partilha em repositórios de AA e o trabalho para a reutilização dos dados. Foram também destacados, os metadados ou a documentação necessária para a contextualização dos dados; os princípios FAIR, bem como a importância das publicações e das políticas que acompanham o contexto da abertura. Este conjunto de itens, julgados como gerais, são basilares para formar uma visão abrangente para a abertura e partilha dos dados de investigação, conforme representado na figura abaixo.

Figura 7

Visão geral dos elementos para a abertura e partilha dos dados em acesso aberto



A proposta de uma visão geral para abertura dos dados de investigação, constituída pelos itens basilares da figura anterior, passa necessariamente pelo seguinte entendimento:

1) Os dados produzidos na investigação, sob os princípios e práticas da CA, necessitam ser abertos devido ao contexto atual da ciência e da comunicação científica, dos financiamentos e das políticas que as recomendam, e às vezes exigem, a abertura dos dados.

2) Os repositórios digitais de AA, são plataformas utilizadas para a partilha dos dados, possibilitando realizar os objetivos declarados pelos princípios e práticas da CA.

3) Através da utilização dos princípios FAIR, o processo para a abertura dos dados, acompanhado por documentação e metadados, pode resultar no uso atual e futuro dos dados por investigadores, proporcionando os benefícios esperados do processo.

Por intermédio de uma visão geral para abertura dos dados de investigação, resultado do referencial teórico aqui trabalhado, os seguintes itens a serem levados em consideração para realizar uma partilha adequada dos dados de investigação, no ambiente da universidade e sob o contexto do CA, podem ser listados e interrelacionados: a) a investigação e o investigador sob os princípios e práticas da CA; b) a comunicação da ciência; c) as políticas e normas orientadoras e fomentadoras; d) as infraestruturas como os repositórios digitais; e) a documentação e os metadados que acompanham os dados; f) os princípios FAIR; e, f) a atenção para uma reutilização dos dados na atualidade e no futuro.

3.6. CONSIDERAÇÕES PARA REALIZAR UMA PARTILHA ADEQUADA DOS DADOS

Segundo Borgman (2018, p.370), “Identifying problems and principles is an essential starting point to address the challenges of the day. Applying these principles to solve these problems is much harder”. Assim, foram muitos os princípios, fatores, características, propósitos, aspectos, vantagens e desvantagens identificados ao longo deste capítulo relacionados com a ciência aberta e a partilha dos dados de investigação, citados com a finalidade em responder à questão complementar inicialmente especificada (Tabela 13).

Por conseguinte, foi trilhado o caminho no sentido de conhecer a importância e as características da partilha em AA, para depois selecionar, organizar e relacionar sob aspectos temáticos, os princípios e as práticas da CA orientados pelo propósito da partilha dos dados

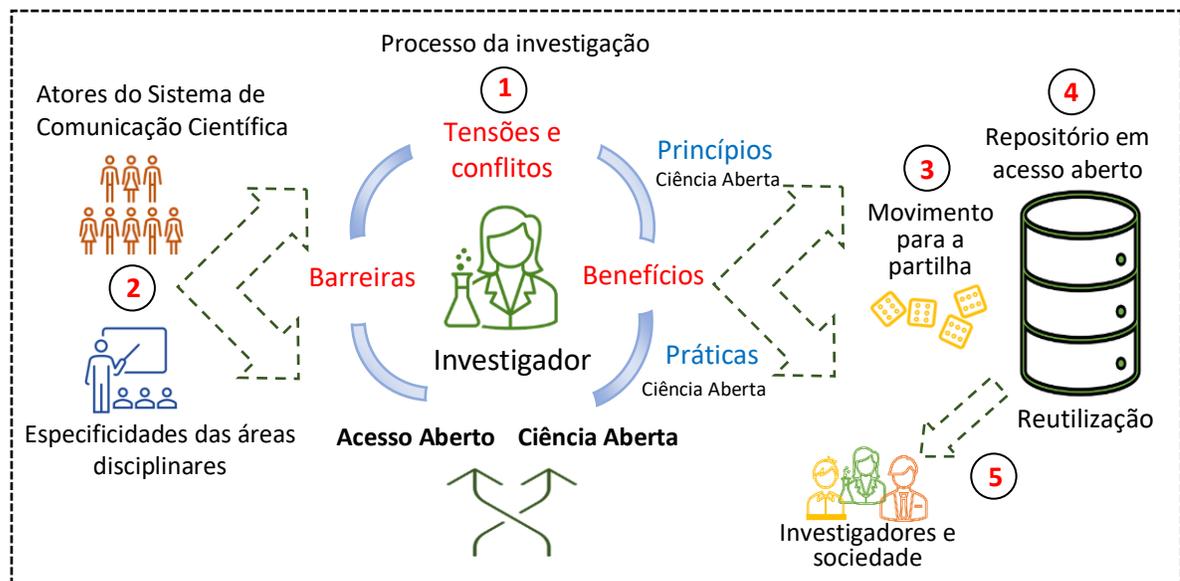
em AA. Em seguida, ressaltou-se o lado cognitivo e contextual do investigador na forma de tensões e conflitos, bem como os atores capazes de influenciá-lo no ato de partilhar os dados. Por fim, destacou-se uma visão geral para a abertura e para a partilha dos dados.

É sob a perspectiva resumida acima que foi aqui apresentado um caminho para responder à questão complementar: o que é uma partilha adequada dos dados? Tal caminho possibilitou a reunião de elementos a serem considerados para realização de uma partilha adequada dos dados de investigação, no ambiente da universidade e sob o contexto da CA.

Nesse sentido, dois produtos resultantes deste capítulo são apresentados: 1) o primeiro é a figura que segue, onde é mostrado um panorama da sintetização do contexto e da relação dos principais elementos aqui referenciados, sob uma ideia processual; e, 2) o segundo é a tabela 14, que tendo por base o item anterior, reúne os elementos que ajudam a responder à questão complementar proposta, ao mesmo tempo que pode ser percebida como o resumo dos itens abordados no presente capítulo.

Figura 8

Contexto e elementos para uma partilha adequada dos dados na universidade



O panorama mostrado na figura 8 que representa o contexto da universidade e os elementos aqui encontrados, a serem considerados em uma partilha adequada dos dados de investigação, são nomeados e explicados conforme a numeração apresentada como segue:

1) **Influências** – são as tensões e conflitos, barreiras ou benefícios, que podem afetar o investigador no processo da investigação, manifestados de maneira cognitiva por questões individuais, contextuais ou institucionais, são pressões sobre o investigador, exercidas de modo a influenciá-lo na decisão de como partilhar, ou mesmo se deve partilhar.

2) **Especificidades e agentes relacionais** – são os elementos contextuais ou institucionais da investigação aqui caracterizados, mas não a estes limitados, como os agentes do sistema de comunicação científica e as especificidades das áreas disciplinares da universidade, que podem interferir em relação ao investigador sobre o ato da partilha.

3) **Movimento para a partilha dos dados** – é o conjunto de ações, que deve ser observado como resultado do uso pertinente dos princípios e práticas da CA, de modo lógico, organizado e relacionados entre si para atingir a finalidade da partilha dos dados por infraestruturas digitais em acesso aberto, como os repositórios de AA, considerando também a comunicação científica, os metadados, os princípios FAIR e as políticas, de modo que a partilha possibilite a reutilização dos dados.

4) **Infraestrutura digital de AA** – é parte do processo de abertura dos dados e resultante do movimento anterior, os dados partilhados pelo investigador permitem serem reutilizados através de uma infraestrutura digital, objetivada pelos princípios e práticas da CA e pelo AA, referenciado geralmente pelos repositórios digitais.

5) **Reutilização dos dados por investigadores e pela sociedade** – é o resultado da partilha dos dados sob os princípios e práticas da CA, de modo a possibilitar o uso atual e futuro dos dados, tanto pelos investigadores quanto de uma forma geral pela sociedade, o que permite resultar no avanço da ciência e na realização de benefícios sociais.

Assumindo como base os elementos e o contexto sintetizados na figura 8, a tabela que segue apresenta de modo mais detalhado o que se deve considerar e as justificativas que ajudam a responder à questão complementar proposta para este capítulo, de modo também a reunir os principais itens de estudo que foram aqui destacados e que servem de base para conceber o que seja uma partilha adequada dos dados de investigação.

Tabela 14

O que considerar para a realização de uma partilha adequada dos dados?

O que considerar	Porque considerar
O acesso aberto	Constitui um dos pilares da ciência aberta que é básico para a partilha dos dados, no sentido da disponibilização, de forma pública, digital, <i>online</i> , irrestrita e gratuita, ou com um menor número de restrições e menor custo possível.
O investigador na investigação local	Embora a partilha em AA possa alcançar objetivos globais, é necessário direcionar a atenção primeiro para a investigação local e nas ações do investigador, pois é destas ações que depende uma partilha adequada dos dados.
A ação da partilha	O modo como é realizada a partilha pode ser resultado das influências sobre o investigador, devendo ser observada para possíveis providências no intento de alcançar uma partilha adequada dos dados.
As formas e os meios da partilha	Considera-se ideal que uma partilha adequada seja feita em repositórios digitais, <i>online</i> , livre para uso, sem ou com mínimas barreiras, respeitando os direitos constituídos e a gratuidade quando possível.
Os princípios e as práticas da ciência aberta	Os fatores positivos e os benefícios justificados pelos princípios e pelas práticas da CA, são elementos orientadores adequados para realização de uma partilha em AA.
A organização relacional por aspectos temáticos	Os princípios e práticas da CA, agrupados em aspectos temáticos, podem orientar de forma adequada a partilha em AA, quando organizados de modo lógico, proposital e com relacionamentos bem definidos.
As tensões e conflitos	As preocupações em nível cognitivo e contextual do investigador, bem como os fatores negativos percebidos para a partilha, podem gerar tensões que contrastados com os fatores positivos podem gerar conflitos, resultando em barreiras para a partilha dos dados.
As barreiras	São motivos que afetam o investigador para não partilhar os dados, como a perda de oportunidade, esforço adicional, desconfiança/plágio, desconhecimento, informações sigilosas, entre outros.
Os benefícios	São as justificativas que afetam o investigador para partilhar os dados, como a acessibilidade; visibilidade; credibilidade; recompensa acadêmica; reconhecimento profissional e altruísmo, entre outros.
As especificidades do contexto	As peculiaridades do ambiente podem influenciar na partilha, como as diferenças disciplinares e o relacionamento com os atores do sistema de comunicação científica no contexto da universidade.
As diferenças disciplinares	O comportamento dos investigadores para a partilha pode variar de acordo com a área disciplinar e as práticas tradicionais na qual estão inseridos.
Os agentes do sistema de comunicação científica	Através de ações ou por sua falta, os investigadores, as editoras, bibliotecas, as agências de fomento e as universidades podem influir na partilha dos dados.
Os atores para abertura dos dados de investigação	A partilha em repositórios digitais deve ser realizada através dos princípios e práticas da CA, pela necessidade da comunicação científica, no cumprimento das políticas, com o uso de metadados e dos princípios FAIR, objetivados para a reutilização atual e futura dos dados de investigação.

Propõe-se também, tendo por base o construto anterior, como deve ser uma partilha adequada dos dados de investigação:

- **Centrada**, de modo a considerar o investigador e suas ações como foco principal;
- **Sociotécnica**, pelo fato de que a tecnologia e os fatores sociais estão imbricados e são indissociáveis do contexto da partilha;
- **Fundamentada**, a fim de que seja direcionada pelos princípios e práticas da CA, de modo coerente e estruturado, na condução da atividade para a partilha em acesso aberto;
- **Organizada**, a fim de realizar uma partilha de forma lógica e proposital visando minimizar a desatenção no ato da partilha;
- **Cognitiva**, para atentar que o investigador pode ser influenciado por fatores individuais e contextuais na realização da partilha;
- **Flexível**, para que as peculiaridades do contexto e do ambiente não sejam ignoradas no processo da partilha;
- **Sistêmica**, com a finalidade de que os atores e as suas interações sejam interrelacionados e influentes em todos os níveis para o objetivo da partilha; e
- **Direcionada**, de forma a intencionalmente utilizar os instrumentos e os processos necessários em uma partilha objetivada para a reutilização dos dados.

O contexto, os elementos e o modo para realizar uma partilha adequada dos dados de investigação apresentados nesta seção, reúne os itens necessários para caracterizar este tipo de partilha objetivada na tese. Neste sentido, o capítulo atual destacou e organizou os elementos encontrados na literatura a serem considerados para tal caracterização, o que resulta na definição realizada e apresentada no capítulo 6 e constante na tabela 23. Cumprindo o segundo objetivo específico do trabalho.

Para finalizar a seção, cabe ressaltar que o processo de investigação, embora aqui citado, não foi alvo de destaque, do mesmo modo as etapas que podem envolver a abertura para a partilha dos dados. Estes itens também são fundamentais para identificar as práticas essenciais para a RDM em uma partilha adequada dos dados de investigação. Assim, o próximo capítulo aborda a temática necessária para o preenchimento destas lacunas.

4. MODELOS DE CICLOS DE VIDA – PRÁTICAS ESSENCIAIS PARA A GESTÃO DOS DADOS

Este capítulo explora principalmente as etapas, os elementos e os processos constantes nos modelos de ciclos de vida³⁵ levantados na literatura científica. São estruturas que representam os fluxos da RDM (*Research Data Management*), da curadoria e dos dados de investigação. Esta exploração visa identificar e descrever as práticas³⁶ essenciais para a RDM que permitam sustentar uma partilha adequada dos dados na universidade. Para isto, os modelos serão estudados de modo a possibilitar: 1) identificar quais sejam as práticas recorrentes; 2) reconhecer os elementos contextuais que envolvem tais práticas; e, 3) sintetizar os itens anteriores para a obtenção das práticas essenciais objetivadas. Portanto, o estudo dos modelos, seus elementos e processos, contribuem para a resposta da questão principal deste trabalho no tocante as práticas essenciais para RDM. De modo específico, o capítulo cumpre o objetivo três da tese e atende à questão complementar: quais são as práticas essenciais para a RDM?

Os modelos que envolvem os fluxos nos ciclos de vida têm sido representados de diferentes maneiras (Möller, 2013). Portanto, identifica-se que as representações aqui examinadas são constituídas por processos, fases ou etapas³⁷, atores, práticas, características e fundamentos referentes aos dados de investigação, mas também à RDM e à curadoria digital. Assim, os modelos podem ter seus fluxos estudados para identificar um contexto necessário e combinar os elementos e as principais práticas envolvidas na RDM para a partilha dos dados, tornando-se útil para respaldar teoricamente recomendações de práticas institucionais, realizada por investigadores ou por outros atores que lidam com os dados, com a sua gestão e curadoria em ambientes como a universidade.

³⁵ O termo “modelos de ciclo de vida” é aqui utilizado indistintamente para as representações dos fluxos da RDM, da curadoria e dos dados de investigação. No entendimento de Borgman (2015), o “ciclo de vida” é um termo comumente utilizado para refletir as mudanças na forma e no caráter dos objetos ao longo do tempo, na origem até sua disposição final, que pode ser a preservação ou a destruição e, no contexto da comunicação científica podem referir-se as fases da investigação, da origem das ideias até a publicação e a disseminação. Os ciclos de vida se cruzam, de modo que diferentes produtos são criados e usados em cada estágio da investigação. A autora afirma que apesar do amplo uso, o termo permanece problemático, pois sugere que a investigação prossegue gradativamente com um começo, meio e fim, ao invés de iterar por muitas etapas.

³⁶ As “práticas” também se referem as técnicas, ferramentas, tecnologias, metodologias, abordagens, iniciativas e instrumentos que possam ser utilizados para apoiar a gestão e a partilha dos dados de investigação.

³⁷ Os termos “fases”, “etapas”, ou “estágios, referente aos ciclos de vida, são aqui usados sem diferenciação.

Para uma melhor compreensão e base para a análise dos modelos explorados, é fundamental realizar uma breve contextualização teórica para a diferenciação entre as práticas de curadoria digital e as práticas da RDM. Isto é necessário devido à importância que o tema e o termo “curadoria” exercem na área, mas também porque a literatura tende a não ser clara em tal diferenciação, sendo que ambos juntamente com o processo de investigação estão representados nos modelos aqui estudados. Portanto, é necessário obter uma visão geral do tema, para clarificar, relacionar e estabelecer limites de atuação entre eles, para então abordar os modelos conforme o objetivo do capítulo.

4.1. PRÁTICAS DE CURADORIA E DE GESTÃO DOS DADOS

As diversas referências às práticas de curadoria digital e a novidade emergente do assunto em torno da RDM criam uma variabilidade de títulos, termos e neologismos no âmbito do tema da curadoria (Tammaro, Sposito & Casarosa, 2019). O termo “curadoria digital” usado na literatura beneficiou-se da ideia originária de curadoria das áreas da biblioteca, museu e ciências biológicas, ou seja, implica na preservação e na manutenção de uma coleção, adicionado valor de conhecimento, acompanhado da descrição do contexto e do histórico (Beagrie, 2006). Porém, com o aumento do volume da informação digital e conseqüentemente dos dados de investigação, a curadoria ganhou outras características e atribuições para possibilitar organizar e gerir tais dados, bem como o material digital.

Em geral, as definições e citações sobre a curadoria digital propõem gerir os dados através de ciclos de vida no processo da investigação. As referências reúnem diversos atores em torno dos termos “gestão” e “curadoria”, porém sem clareza de distinção, apenas fazem menção, e não poucas vezes igualam as práticas envolvidas na RDM com as práticas de curadoria (Borgman, 2010; Heidorn, 2008; Lord & Macdonald, 2003; Weidner & Alemneh, 2013). O problema é que com esta falta de distinção, a gestão dos dados pode ser realizada sem a reflexão necessária, de modo a resultar em práticas inadequadas para a RDM. Assim, é necessário verificar as diferenciações, convergências e limites entre as práticas de curadoria digital e da RDM no processo de investigação, considerando o ciclo de vida dos dados. O entendimento desta perspectiva contribui também para obter uma visão geral do tema, de modo a melhor interpretar os elementos envolvidos nos modelos aqui estudados.

É a partir do significado de curadoria digital que serão estabelecidas as diferenciações, convergências e limites com a RDM. Assim, em uma definição largamente referenciada por autores da área, o Digital Curation Center (DCC) afirma que: a “Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle [...] enhances the long-term value of existing data by making it available for further high-quality research” (Digital Curation Center, 2021a, para.3). No mesmo sentido, Tenopir et al. (2013), indicam que a curadoria digital, além das etapas referenciadas pelo DCC, também abarca as fases de seleção e arquivamento dos dados.

No estudo de Kouper (2016), a autora acrescenta mais fases no processo de curadoria, afirmando que, apesar de existirem algumas diferenças nas definições e tentativas de concentração em um aspecto ou outro, de modo geral, a curadoria digital abrange todas as atividades relacionadas ao cuidado de recursos digitais em ações de criação, seleção, avaliação, descrição, classificação, anotação, armazenamento, preservação, acesso, descoberta, transformação, descarte, disseminação e a divulgação dos dados. Portanto, por esta atuação alargada, a curadoria digital pode abranger diversas atividades do tratamento dos dados, da criação até a reutilização, incluindo os repositórios e a sua gestão (Abbott, 2008; Constantopoulos & Dallas, 2010; Lord et al., 2004).

Como referido na definição do DCC, as práticas de curadoria são representadas nos modelos de ciclos de vida. Neste sentido, são várias as definições que abordam, destacam e direcionam tais práticas nas etapas dos modelos (Beagrie, 2006; Dempsey, 2004; Higgins, 2008; Lavoie, 2010; Lee & Tibbo, 2007; Winget, et al., 2009). Geralmente o processo de investigação também está representado nos modelos de ciclo de vida com ênfase nos dados, pois é na investigação que os dados são gerados e a curadoria é iniciada e padronizada. Recomenda-se, inclusive, que a curadoria digital seja instaurada o mais cedo possível no processo de investigação (Ball, 2012; Tibbo, 2015; Cox & Verbaan, 2018).

No âmbito do processo de investigação e da informação digital, Humphrey (2006), assim explica a importância dos modelos de ciclo de vida:

Life cycle models are shaping the way we study digital information processes. These models represent the life course of a larger system, such as the research process, through a series of sequentially related stages or phases in which information is produced or manipulated (p.1).

A RDM também remete as suas práticas aos modelos de ciclos de vida, imbricados com os ciclos de vida dos dados (Cox & Pinfield, 2013; Whyte & Tedds, 2011). Tais modelos decompõem os fenômenos da RDM em uma série de tarefas, que são direcionadas a diferentes papéis ou atores, levando em consideração diversas questões, como as legais, éticas, estruturais e relacionamentos institucionais e de governança (Weber & Kranzlmüller, 2019).

As definições para a RDM, como a de Cox e Pinfield (2013), associam tal gestão ao ciclo de vida dos dados, em uma atividade que envolve o projeto e a criação de dados, armazenamento, segurança, preservação, recuperação, partilha, reutilização, questões éticas e legais, entre outros, verifica-se a semelhança com as práticas de curadoria. Corti et al. (2011), também sugerem que a RDM deve abordar questões sobre os dados gerados na investigação, bem como metadados, padrões, armazenamento, *backup*, segurança e partilha, direitos autorais, propriedade intelectual, recursos e custos associados e funções e responsabilidades. Cox e Verbaan (2018) descreveram a RDM como um processo que inclui criar, encontrar, organizar, armazenar, partilhar e preservar dados de investigação em um ciclo de vida.

Assim, pode-se perceber que as práticas de curadoria e para a RDM normalmente estão representadas através de ciclos de vida no processo de investigação. Tais representações emergem da necessidade de organizar e obter os benefícios proporcionados por uma estrutura que permite contextualizar momentos, características e requisitos, bem como os fatores que acompanham as transformações dos dados nas etapas, como afirma Carlson (2014) ao citar os benefícios da aplicação dos modelos de ciclos de vida neste contexto:

Applying life cycle models to support services for managing research data has several benefits [...] from its inception to its use and completion, research data will likely undergo multiple transformations in its format, application, use and perhaps even its purpose. Through identifying and naming the transformations that data will undergo as stages in a larger life cycle, organizations can better target their services [...] (pp.63; 65).

Segundo Shepanek (2011), o problema é que tradicionalmente as discussões têm sido estreitamente focadas no ciclo de vida dos dados. Esta perspectiva “miope”, segundo o autor, perde o contexto mais amplo dentro do qual os dados são criados, gerenciados, preservados, podendo levar a abordagens inadequadas. Isto ratifica a importância da observação dos modelos de ciclos de vida de modo abrangente, juntamente com os dados no processo de investigação e com as práticas de curadoria e para a RDM, pois além da

possibilidade de identificação e sistematização de tais práticas, muitas decisões sobre os dados são tomadas em cada etapa do ciclo de vida (Sant’Ana, 2016; Wallis et al., 2008).

Para o estabelecimento de limites, diferenças e convergências no ambiente da universidade, tendo por consideração um contexto onde tanto as definições para a curadoria quanto as de RDM, remetem suas práticas às etapas do ciclo de vida dos dados no processo de investigação, uma possível diferenciação se dá no fato de que enquanto a curadoria está preocupada com ações práticas referentes aos dados, à RDM inclui estas ações em seu escopo e abrange outras atividades com várias partes interessadas, tanto dentro quanto fora da comunidade universitária (Flores et al., 2015). Assim, pode-se inferir que a RDM contém a curadoria e a excede, nas relações com atores internos e externos à universidade.

Uma convergência entre curadoria digital e a RDM, também possível de ser estabelecida, está na percepção de que os termos usados para descrever as etapas e as práticas de curadoria digital, sob o ciclo de vida dos dados, são comuns a ambos. Ou seja, termos atribuídos as práticas de curadoria, como “disseminação”, “reutilização”, “armazenamento”, “preservação”, “acesso”, “descoberta” (Kouper, 2016; Tenopir et al., 2013), têm por origem os princípios e práticas da CA sob o contexto do AA. Estas práticas, igualmente atribuídas à RDM, são da mesma forma originárias da CA sob o AA. Assim, pode-se entender que tais princípios e práticas da CA e do AA, na perspectiva da RDM e da curadoria, são elementos que os unem e os direcionam ao mesmo objetivo.

Aqui, tanto a curadoria quanto a RDM têm em seu eixo o processo de investigação, que, por sua vez, remete a um conjunto de procedimentos intelectuais e técnicos utilizados para atingir um conhecimento. No processo de investigação, existe uma diversidade de métodos, que são determinados pelo tipo de objeto a investigar e pelas proposições a descobrir, viabilizado por etapas pré-determinadas (Gil, 1999). Os dados são parte do processo de investigação, em práticas como a recolha, tabulação, análise ou na discussão dos dados, mas para isso é preciso armazená-los, ou seja, é necessário de alguma forma um tratamento para o uso dos dados. Segundo Eco (1977), o investigador coloca suas ideias em ordem no intuito de organizar os dados obtidos. Portanto, o processo de investigação também contém práticas de curadoria e de RDM para chegar aos resultados da investigação.

Em outra perspectiva pode-se entender que as práticas de curadoria e para a RDM estão contidas no processo de investigação, mas também excedem tal processo, pois tratam de abordagens que remetem a ações posteriores como, por exemplo, atividades em repositórios. No entanto, estas práticas também precedem à investigação a exemplo da necessidade de um plano para a curadoria dos dados e nas relações institucionais para a realização da RDM. Assim, o escopo de atuação das ações da curadoria e para a RDM precedem, estão contidas e excedem o processo de investigação, por se referirem a práticas necessárias de serem realizadas antes mesmo da entrada no processo da investigação e após sua finalização, estabelecendo limites, relações fundamentais e a abrangência da atividade.

Para organizar as inferências conceituais anteriores, a figura 9 propõe uma visualização geral das diferenciações, convergências e limites entre as práticas de curadoria digital e da RDM no processo de investigação, considerando o ciclo de vida dos dados. Nesta proposição, os atores referenciados no capítulo anterior são mostrados na figura abaixo sem destaque, pois como o objetivo neste momento não são os atores, estes servem como indicação de contexto para a gestão e partilha dos dados no ambiente da universidade.

Figura 9

Visão geral das convergências, diferenciações e limites entre as práticas de curadoria e para a RDM no processo de investigação, considerando o ciclo de vida dos dados



A compreensão das diferenciações, convergências e limites propostos, podem ser resumidas na seguinte interpretação da figura acima:

1) As práticas para a RDM abrangem todo o processo de investigação, bem como as ações de curadoria digital e o ciclo de vida dos dados, porém estas práticas não estão

limitadas a tal abrangência, dado que a RDM apresenta práticas que precedem e excedem o processo de investigação;

2) O processo de investigação contém as práticas de curadoria digital especificadas nos ciclos de vida dos dados, sendo também parte fundamental e central para que as práticas para a RDM possam ser realizadas;

3) As práticas de curadoria digital fazem parte do processo de investigação, mas o transcende em ações específicas, são parte da RDM com práticas direcionadas pelo ciclo de vida dos dados, tendo como intersecção os princípios e as práticas da CA e o AA;

4) A diferenciação entre a RDM e a curadoria está nos relacionamentos internos e externos à universidade, que ultrapassam os limites das ações no tratamento com os dados;

5) O ciclo de vida dos dados, representa as práticas de curadoria relacionadas à RDM e está contido, mas não limitado, ao processo de investigação; e

6) Todas as práticas têm como sustentação os princípios da CA e o acesso aberto.

É possível perceber que é necessário associar as práticas para a RDM e da curadoria ao ambiente, para identificar as diferenciações, convergências e limites entre elas. Mas também, e de modo geral, percebe-se que a curadoria e a RDM têm objetivos semelhantes em relação às práticas para o tratamento dos dados, de modo a parecer tênue a linha divisória entre as duas. Daí podem-se entender certas sobreposições e explicar o tratamento indiferenciado entre eles, mas da mesma forma também possibilita o questionamento do uso de dois termos para a mesma finalidade ou finalidades muito semelhantes. No entanto, um fator determinante citado, mas não explorado neste momento, que bem permite realçar e diferenciar conceitualmente os limites entre as práticas da RDM e da curadoria, são os atores com atribuições nos relacionamentos institucionais.

Cabe ressaltar que a visão geral aqui apresentada, tem aplicação específica para exemplificar limites globais entre as práticas de curadoria digital e da gestão dos dados, no ambiente do processo de investigação e no ciclo de vida dos dados, isto com o claro objetivo de obter uma melhor compreensão e análise dos elementos e processos envolvidos nos diversos modelos estudados na próxima seção. Assim, esta breve reflexão não teve a

preocupação em definir de modo determinante os limites entre a RDM e a curadoria, nem abordar todas as possibilidades, mas propiciar um quadro mais abrangente e minimamente esclarecedor para ajudar a identificar as práticas essenciais para a RDM objetivada na sustentação da partilha dos dados de investigação.

Portanto, esta seção verificou de modo teórico, as diferenciações, convergências e limites entre as práticas de curadoria digital e para a RDM no processo de investigação, considerando o ciclo de vida dos dados. O resultado contribuiu para a proposição de uma visão geral do que foi verificado, de modo a melhor interpretar os elementos e processos envolvidos e constituintes dos modelos estudados a seguir.

4.2. MODELOS DE CICLOS DE VIDA

Os modelos que representam os ciclos de vida têm se transformado em importantes ferramentas no trabalho científico, o que se torna cada vez mais pertinente com a ênfase crescente na partilha dos dados na investigação (Corti et al., 2014). Estes modelos são estruturas relevantes, pois tanto ajudam na partilha dos dados pelos atores envolvidos, quanto na identificação e visualização das etapas no processo de investigação, facilitando as práticas para a análise dos dados (Tammaro, Sposito & Casarosa, 2019). São, portanto, ferramentas úteis para o planejamento e desenvolvimento dos serviços para a RDM, mas também na preservação e retenção dos dados em longo prazo (Carlson, 2014; Higgins, 2008).

O estudo das práticas para a RDM através dos modelos de ciclos de vida, além de servir de base para compreensão das fases, relacionamento entre os processos e os fatores envolvidos que podem ser aplicados à gestão e a partilha dos dados, são essenciais para a identificação e otimização dos dados, mas também para conhecer os recursos disponíveis, obter o melhor proveito destes e as características que propiciam o equilíbrio entre os atores incluídos no processo para a RDM (Möller, 2013; Sant'Ana, 2013). Weber e Kranzlmülle (2019), apontam a importância dos modelos de ciclos de vida para a RDM assim:

A large number of services, tools, protocols, best practices, and policies have been created and are currently competing for adoption. This state [...] leads itself to a research question: How can we describe, explain, assess, and maybe even predict phenomena in research data management? Of what nature is the interaction between researchers and other professionals? The most common answer to this question is to model the phenomena of research data management along a lifecycle (p. 75).

Como declarado no início do capítulo, o estudo dos modelos neste trabalho tem o objetivo de identificar e descrever as práticas essenciais para a RDM, de modo a permitir sustentar uma partilha adequada dos dados, no ambiente da universidade. Para tal, os modelos aqui apresentados serão analisados destacando primordialmente três grupos de ações, quais sejam: 1) a contextualização e a caracterização dos modelos; 2) a análise das práticas, processos e os elementos que possam ser úteis para a RDM em uma partilha dos dados; e, 3) a seleção e a sintetização das etapas dos modelos para que as práticas essenciais sejam identificadas, caracterizadas e apresentadas.

É importante explicar que esta abordagem obviamente não contempla todos os modelos de ciclos de vida, pois há uma grande quantidade de modelos semelhantes e etapas repetidas, decorrentes das derivações que acontecem uns dos outros. Assim, com os métodos empregados conforme especificados na metodologia desta tese, chegou-se, portanto, a nove modelos finais. Os modelos foram selecionados através da pesquisa bibliográfica, mas também no estudo de artigos que já contivessem uma grande quantidade de modelos³⁸ de ciclos de vida analisados. Usando assim, a pesquisa bibliográfica e o referencial teórico direcionado pela literatura da temática.

Em decorrência do exposto, foram selecionados os modelos percebidos como mais relevantes, sendo obtidos com os parâmetros que permitem a aplicação no contexto da RDM e da partilha dos dados. Os parâmetros empreendidos foram os seguintes: a) deve haver citação recorrente na literatura; b) ser dedicado aos dados de investigação e/ou a gestão e/ou a curadoria digital; c) conter o ciclo de vida da investigação; d) poder ser identificados através de padrões e/ou práticas; e, d) ter peculiaridades distintivas, de modo a permitir não obter somente modelos semelhantes. Assim, a tabela que segue lista os nove modelos que representam os fluxos da RDM, da curadoria e dos dados de investigação selecionados, juntamente com as características que foram eletivas para que fizessem parte deste estudo.

³⁸ Em contexto e objetivos diferentes, os autores dos artigos selecionados analisaram e revisaram os seguintes quantitativos de modelos de ciclos de vida - Ball (2012) nove modelos; Comitê de Satélites de Observação da Terra (CEOS, 2012) quarenta e quatro modelos; Cox e Tam (2018) nove modelos; Gupta e Müller-Birn (2018) cinco modelos; Weber e Kranzlmüller (2019) resumem outros noventa modelos. Perrier et al. (2017), fornecem uma revisão de escopo de 301 artigos e 10 documentos complementares que discutem as práticas de RDM em instituições acadêmicas entre 1995 e 2016. Estes artigos de revisão não se limitam, mas incluem de forma principal as publicações que discutem os modelos de ciclos de vida no contexto aqui estudado.

Tabela 15

Modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação

Ciclo de vida	Instituição/autor de referência	Características eletivas
DCC	Digital Curation Centre (Higgins, 2008)	Centrado na curadoria digital, é uma importante referência na área, apresenta-se com 3 grupos de práticas, distribuídos em 15 ações principais para a RDM, passíveis de detalhamento e expansão.
DDI	Data Documentation Initiative (Structural Reform Group, 2004)	É um modelo matriz, básico para concepção de estruturas com ênfase na investigação e no tratamento dos dados, adotado por diversas instituições e universidades para o uso com a RDM.
e-Science	(Humphrey, 2006)	Derivação do modelo DDI, destaca a e-Science e aborda a perda de informação. Este modelo é citado na literatura de forma recorrente.
UVa	Virginia University (University of Virginia Lybrary, 2022)	É outra derivação do modelo DDI, apresenta de forma organizada a percepção de uma biblioteca universitária para o ciclo de vida dos dados e para o desenvolvimento da RDM.
DataONE	Data Observation Network for Earth (Strasser et al, 2012)	Diferente dos outros modelos, a ênfase do modelo DataONE é quase que exclusivamente no tratamento dos dados e na RDM, com menor ênfase na investigação.
Jisc	Joint Information Systems Committee (Jisc, 2021)	O modelo provém de uma importante organização que trabalha com as universidades e a investigação no Reino Unido, é direcionado à curadoria digital e à RDM, porém mais simples que o modelo do DCC e com um abrangente detalhamento dos itens a serem observados para cada etapa do ciclo de vida.
UK Data Archive	University Of Essex (Corti et al., 2014)	É um modelo circular com poucos estágios, a ênfase encontra-se na associação dos dados com o processo de investigação.
CENS	Center For Embedded Networked Sensing (Wallis et al., 2008)	O modelo do ciclo de vida do CENS contou em seu grupo de autores com uma das principais investigadoras da área, Christine Borgman, o grupo também constrói o modelo através de investigação em uma área heterogênea.
IWGDD	Interagency Working Group on Digital Data (IWGDD, 2009)	Apresenta um modelo circular que aborda aspectos para além do processo da investigação e da curadoria, como as políticas e habilidades profissionais em torno da RDM.

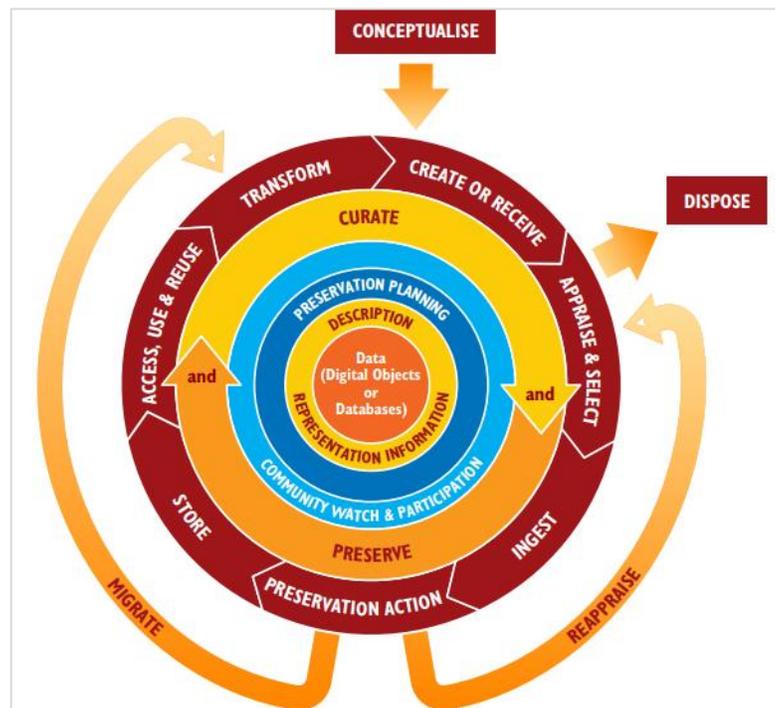
Consoante com as ações e os objetivos propostos, faz-se importante assinalar que na análise dos nove modelos, não há comparação direta entre eles e nem uma avaliação ampla de modo crítico, mas uma análise descritiva que resulta na reunião dos elementos que permitem identificar as práticas essenciais para a RDM, objetivadas para uma partilha adequada dos dados. A análise dos modelos seguem na mesma ordem da tabela acima.

4.2.1. DCC - DIGITAL CURATION CENTRE

O modelo do ciclo de vida do *Digital Curation Centre* – DCC, foi apresentado formalmente em 2008 no Reino Unido através do trabalho de Sara Higgins. Este modelo é bastante influente na área da curadoria dos dados, visto que até 2020 continuava a ser o artigo mais procurado do *International Journal of Digital Curation*³⁹ (Choudhury, Huang & Palmer, 2020). Fornecendo uma visão de alto nível dos estágios do ciclo de vida da curadoria, gerenciamento e preservação de dados, o modelo é intencionalmente genérico, indicativo em vez de exaustivo, adaptável e aplicável em diferentes domínios e níveis de granularidade. Desenvolvido como uma ferramenta específica, pode ser usado em conjunto com as normas, para planejar atividades de curadoria, preservação e ações dos produtores dos dados, curadores e outros usuários (Digital Curation Centre, 2021a; Higgins, 2008).

Figura 10

Modelo do ciclo de vida DCC – Digital Curation Centre



Nota. Elaborado por Higgins (2008).

³⁹ O *International Journal of Digital Curation* (IJDC) é um periódico eletrônico de grande repercussão acadêmica, dedicado ao avanço da curadoria digital através de artigos científicos e notícias sobre curadoria de objetos digitais em uma ampla gama de setores e assuntos relacionados ao tema.

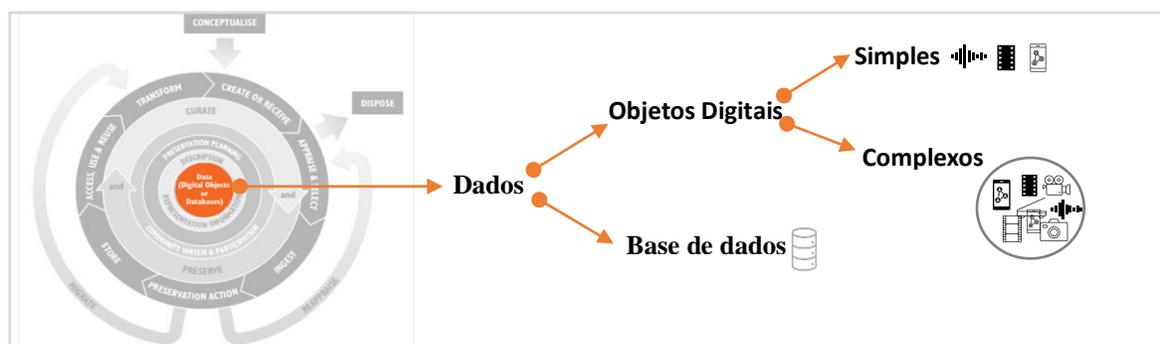
A abordagem do modelo do DCC visa garantir que as etapas essenciais de curadoria sejam planejadas e implementadas de forma ideal. No entanto, a sua operacionalização, dependerá da necessidade do contexto da organização ou domínio, permitindo a entrada para uso em qualquer estágio do ciclo de vida. Assim, o modelo é considerado adaptável para vários propósitos. Em relação à RDM, segundo o DCC, a estrutura ajuda a definir os fluxos de trabalho, funções e responsabilidades a ela associadas e oferecer suporte em uma abordagem holística para o desenvolvimento e a otimização da infraestrutura, podendo ser usado para ajudar as organizações a mapear as atividades da RDM e a fornecer suporte funcionais e operacionais (Digital Curation Centre, 2021; Higgins, 2008).

Ao observar o modelo (Figura 10), ao centro encontra-se o dado digital, seguido por três grupos de práticas. Mais especificamente, os três grupos são compostos por: 1º) quatro ações a serem realizadas durante todo o ciclo de vida; 2º) oito ações sequenciais; e, 3º) três ações ocasionais. Segue abaixo, a caracterização do dado na camada central e dos três grupos citados com suas respectivas práticas de curadoria (Figura 11).

Data (Dado: objeto digital ou base de dados) – é o objeto digital central da curadoria. O DCC divide o objeto digital como simples ou complexos. Os simples, são os arquivos textuais ou outras mídias digitais como som e imagens, junto com seus identificadores e metadados relacionados. Os objetos complexos, são caracterizados por objetos digitais combinados. A base de dados constante no modelo, refere-se as coleções estruturadas de registros ou dados armazenados em um sistema computacional.

Figura 11

Os dados como objeto digital central da curadoria



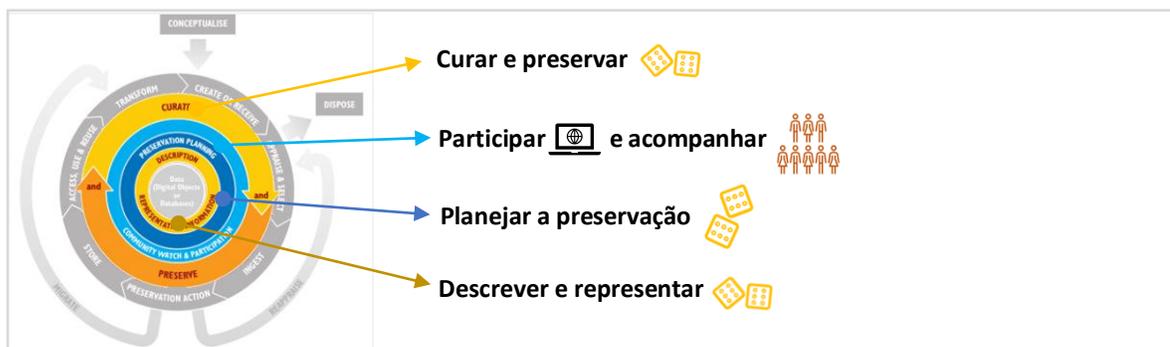
Nota. Adaptado de Higgins (2008).

É interessante perceber que o dado como objeto central da curadoria, está associado as mídias digitais nomeadas como simples ou complexas, e que, embora se possa questionar a definição de tais mídias para a atualidade, a ênfase está no gerenciamento do binário ou digital, e na centralidade das bases de dados. A referência à base ou banco de dados remete a ideia dos repositórios com seus relacionamentos e práticas necessárias para a ação da partilha e da RDM. Assim, nesta parte do modelo, identifica-se a importância dos dados digitais, mas também das ferramentas que podem manipular tais dados, a exemplo dos repositórios, como elemento fundamental para a curadoria.

O primeiro grupo de práticas - **ações para todo o ciclo de vida** (Figura 12), são caracterizados por quatro ações de curadoria que devem ser realizadas durante todo o ciclo de vida do objeto digital, tais práticas são as seguintes:

Figura 12

Primeiro grupo de ações, com 4 práticas para todo o ciclo de vida



Nota. Adaptado de Higgins (2008).

1) **Curate and preserve** (Curar e preservar): realiza ações para a gestão, planejadas de modo a promover a curadoria e a preservação dos dados por todo o ciclo de vida;

2) **Community Watch and Participation** (Participar e acompanhar): monitora as atividades das comunidades envolvidas com a curadoria e a participação do desenvolvimento de ferramentas, padrões e *softwares* para serem compartilhados;

3) **Preservation Planning** (Planejar a preservação): elabora planos de preservação para todas as ações ao longo do ciclo de vida da curadoria, isto inclui planos de gestão de dados e administração direcionadas as ações com os dados e outros objetos digitais; e

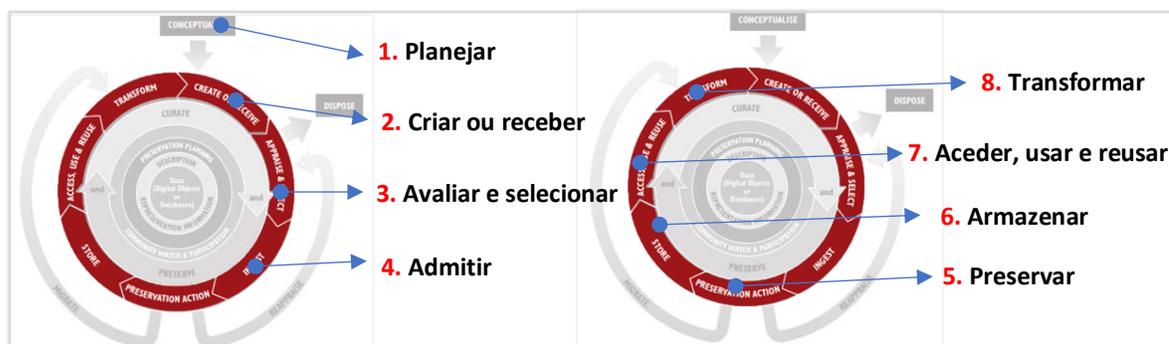
4) *Description and Representation Information* (Descrever e representar a informação): atribui os tipos de metadados pertinentes e necessários para a compreensão contextual e apresentação do dado digital, usando padrões apropriados para garantir a descrição e o controle em longo prazo.

Ao observar este primeiro grupo de práticas, pode-se perceber uma forte ênfase em ações que parecem distintas, mas que fundamentam e são basilares à curadoria no tocante a preservação, administração e participação comunitária. Assim, pode-se perceber que este grupo de práticas destaca além da preservação, o planejamento, o envolvimento com a comunidade, a tecnologia e os metadados. Ou seja, são princípios e práticas da CA a serem realizados durante todo o ciclo de vida da curadoria digital.

O segundo grupo de práticas - **ações sequenciais** (Figura 13), é constituído por oito ações realizadas continuamente, sempre que um novo objeto digital precisa ser curado. O intuito é que as ações sejam realizadas sequencialmente para o sucesso da curadoria:

Figura 13

Segundo grupo de ações, com 8 práticas sequenciais para curadoria



Nota. Adaptado de Higgins (2008).

1) *Conceptualise* (Planejar): é o ato de conceber um plano para as atividades, de modo a permitir a geração, coleta e criação dos dados, mas também o plano envolve os métodos de captura e as opções de armazenamento;

2) *Create or Receive* (Criar ou receber): são os dados gerados e registrados originalmente pelos investigadores ou, coletados em outras fontes como os centros de dados, arquivos ou repositórios, aqui deve-se atribuir os metadados apropriados, suficientes e necessários para caracterizar e serem utilizados no contexto;

3) **Appraise and Select** (Avaliar e selecionar): nesta etapa é necessário avaliar os dados e os selecionar para sua preservação e curadoria em longo prazo. As ações devem ser aderentes às orientações documentadas, políticas e exigências legais. Cabe observar que nesta etapa os dados podem ser descartados, eliminados ou admitidos para uso ou reuso;

4) **Ingest** (Arquivar): no caso do objeto digital ser selecionado para a curadoria e destinado a admissão, este momento é considerado ideal para acontecer a transferência dos dados para os arquivos, repositórios, centros de dados ou outros custodiantes;

5) **Preservation Action** (Ações para preservar): possui a finalidade de tratar os dados para que permaneçam autênticos, confiáveis e utilizáveis, mantendo a integridade e garantindo a preservação e retenção em longo prazo. São ações de limpeza dos dados, validação, atribuição de metadados e garantias das estruturas para os dados ou formatos de arquivo aceitáveis. Nota-se aqui, que os dados podem retornar para reavaliação e serem migrados para formatos diferentes;

6) **Store** (Armazenar): é o armazenamento dos dados de maneira segura, respeitando padrões relevantes. As ações a fim de permitir garantias em longo prazo, podem estar associadas a esta etapa, como a manutenção, atualização e *backups* em repositórios;

7) **Access, Use and Reuse** (Acessar, usar e reusar): nesta fase é certificado que os dados estejam acessíveis para uso e reuso, ou seja, fiquem disponíveis para a comunidade envolvida na curadoria e para os usuários interessados no uso e na reutilização dos dados; e

8) **Transform** (Transformar): é a prática de modificar os dados, pode ser por migração para um formato diferente, ou por seleção ou consulta, possibilitando criar resultados recém-derivados, ou seja, é a criação de novos dados, derivados de outros já existentes, com a finalidade da reutilização e a consequente criação de algo novo.

As ações sequenciais são práticas de curadoria que estão ligadas ao processo de investigação no ciclo de vida. Este conjunto de ações consecutivas são essenciais, pois além de normalmente estarem associadas ao processo de investigação, relaciona-se com as outras camadas e fluxos do modelo, conforme pode ser percebido nas definições das seguintes etapas: avaliar e selecionar; arquivar; preservar; e, armazenar, e estas, podem também relacionar-se com o fluxo externo, através das ações de descarte, reavaliação e migração

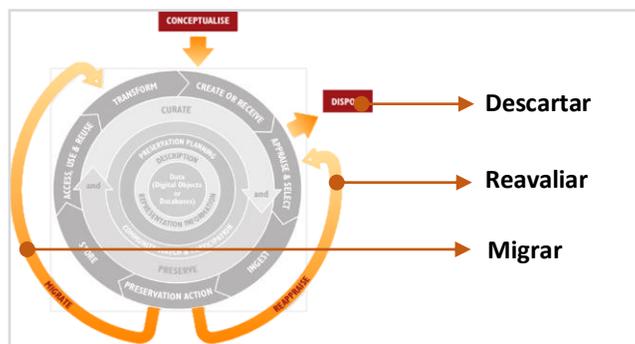
(Figura 14), e ainda, estabelecer relação com a camada anterior referente aos dados (Figura 11), que permite o uso das bases de dados ou dos repositórios.

Não obstante, pode-se questionar se a ordem sequencial e as relações com as outras camadas são as mais adequadas. É sob tal questionamento que o modelo proposto por esta investigação pode adaptar as práticas sequenciais e o processo de investigação, tanto em sua ordem, quanto nos possíveis fluxos relacionais e nas outras partes ou camadas do modelo.

O terceiro grupo de práticas - **Ações ocasionais de curadoria** (Figura 14), é formado por três ações a serem aplicadas conforme as circunstâncias exigirem. A finalidade é gerir o objeto digital, de forma a retirar ou disponibilizar o dado em formato compatível e de acordo com as necessidades apresentadas. Os objetivos e as práticas são as seguintes:

Figura 14

Terceiro grupo de ações com 3 práticas ocasionais de curadoria



Nota. Adaptado de Higgins (2008).

1) **Dispose** (Descartar): refere-se aos dados que não foram selecionados para curadoria e preservação em longo prazo, de acordo com as políticas, orientações ou requisitos legais. Tais dados podem ser descartados ao serem transferidos para outro custodiante. A natureza dos dados pode também exigir uma destruição segura;

2) **Reappraise** (Reavaliar): são os dados que falharam nos procedimentos de validação e passíveis de uma avaliação posterior e nova seleção; e

3) **Migrate** (Migrar): pode-se migrar os dados para um formato diferente, para adaptar ao ambiente de armazenamento ou para garantir a imunidade dos dados, no tocante a obsolescência de *hardware* ou *software*.

As ações ocasionais tem por objetivo permitir que periodicamente os objetos digitais sejam reavaliados (*Reappraise*) para que sua destinação final seja preservada ou descartada (*Dispose*). Essas ações esporádicas servem também para observar se o objeto digital está em um suporte ou formato obsoleto, principalmente nas questões concernentes à informática. A depender desta análise, pode-se passar pelo processo de migração (*Migrate*), assim o acesso ao objeto digital estará preservado. Percebe-se nesta parte do modelo que há uma relação direta com o grupo de ações sequenciais. A ideia de ações ocasionais, que sejam fundamentais para a RDM, a serem utilizadas conforme as necessidades exigirem, são fundamentos pertinentes para o uso do modelo aqui resultante.

Em estudo realizado por Choudhury, Huang e Palmer (2020) sobre a necessidade de análise e atualização do modelo do DCC, com a participação de profissionais das áreas de: informática, biblioteca, preservação digital, ética de dados, investigadores, especialistas em ciência de dados, acadêmicos e curadores de dados, foi realizada entrevistas no *Digital Curation Center*, observando fatores importantes que interessam a esta tese. Os autores apontam uma vertente para a proeminência de dados acionáveis por máquina, bem como a importância do aprendizado de máquina para o processamento e análise dos dados. Também foi apontado o crescimento do uso de algoritmos e fluxos de trabalho de investigação integrados por *softwares*. Embora o assunto não seja alvo específico da análise do modelo, pode ser um elemento a ser considerado e passível de observação para o modelo resultante deste trabalho. Por outro lado, os autores do estudo, também confirmaram o valor contínuo do modelo do DCC e que todos os seus elementos ainda permanecem relevantes.

Esta seção apresentou o modelo do DCC de forma mais detalhada que os outros modelos que serão aqui apresentados, pois o modelo do DCC vem sendo usado como uma ferramenta importante para abertura, curadoria e partilha dos dados de investigação, apresentando-se fundamentada em princípios e práticas da CA e do AA que podem ser usados em uma RDM organizacional. Esta abrangência, como preconizado pelo DCC, requer uma adaptabilidade e flexibilidade ao ambiente de uso, para que o modelo, ou partes dele, possa ser implantado na RDM. No entanto, as práticas citadas e organizadas nos três grupos de ações não podem ser realizadas sem a colaboração e as responsabilidades das pessoas e comunidades que influenciam e atuam nas etapas do ciclo do modelo. Assim, ao

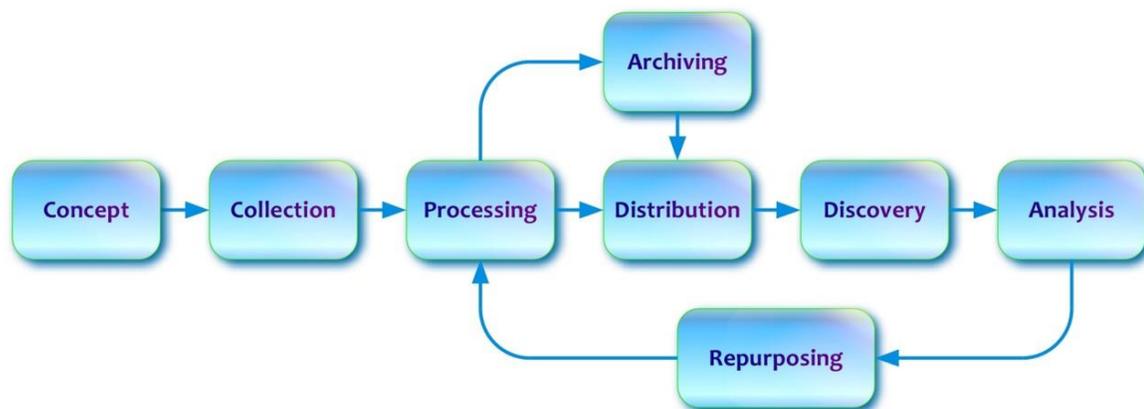
selecionar as práticas essenciais, tais responsabilidades podem constar no modelo resultante. Destaca-se também que a abrangência das práticas, associada aos atores, de modo articulados no modelo do DCC, coaduna com a relevância dos objetivos desta tese.

4.2.2. DDI - DATA DOCUMENTATION INITIATIVE

Estabelecida em 2003, a *Data Documentation Initiative Alliance* (DDI Alliance) é uma colaboração internacional composta por instituições acadêmicas e organizações que trabalham com os dados e a investigação. A DDI Alliance dedica-se a estabelecer padrões de metadados e produtos semânticos para descrever os dados oriundos das ciências sociais e das atividades humanas, tendo por base os métodos observacionais. Em 2004, o modelo de ciclo de vida DDI foi elaborado para documentar e gerenciar os dados por todo o processo da investigação, em seu ferramental de apoio, conta com o *software* livre DDI-*Lifecycle Specification*, que compreende os módulos correspondentes as etapas especificadas no modelo (Figura 15) (Data Documentation Initiative, 2021; Structural Reform Group, 2004).

Figura 15

Modelo de ciclo de vida DDI – Data Documentation Initiative



Nota. Elaborado por Data Documentation Initiative (2021).

O modelo possui oito etapas nas quais ocorrem as práticas em relação aos dados de investigação. A estrutura modelar é definida pela DDI como circular, embora mostre-se em parte linear, com um caminho alternativo e um *loop*. As ações a serem desenvolvidas nas oito etapas podem ser entendidas da seguinte forma (Ball, 2012):

1) **Concept** (Conceito do estudo): é a fase em que a investigação é projetada. O modelo inclui a escolha da questão de investigação e a metodologia para a coleta dos dados, mas também o planeamento para informar como os dados serão processados, analisados e utilizados, bem como o seu formato. Os investigadores definem nesta fase as relações necessárias que existirão em relação aos dados produzidos na investigação;

2) **Collection** (Colecção dos dados): são os dados coletados que incluem como fontes da investigação as estatísticas ou coleções baseadas na *Web* e registros variados. Indica-se que as fontes de dados primários e secundários devem ser claramente distinguidas;

3) **Processing** (Processamento dos dados): uma vez que os dados de entrada são coletados, ocorre o processamento e a análise para produzir os dados de saída, que devem ser registrados em uma forma legível por máquina ou orientada para o ser humano;

4) **Archiving** (Arquivamento dos dados): os dados de saída são arquivados a fim de garantir o acesso em longo prazo, em vez de serem mantidos pelos investigadores. O arquivamento preserva os dados (e metadados) e agrega valor ao longo do tempo;

5) **Distribution** (Distribuição dos dados): o acesso aos dados de saída pode acontecer de duas formas, diretamente, após a etapa de processamento, ou após a etapa de arquivamento, por meio de uma biblioteca ou arquivo de dados;

6) **Discovery** (Descoberta dos dados): para que os usuários descubram os dados, a divulgação pode acontecer por meio de livros em periódicos ou serviços *online*;

7) **Analysis** (Análise de dados): após descobertos, os dados podem ser analisados dentro dos limites do conceito do estudo; e

8) **Repurposing** (Reaproveitamento): decorrente da análise, os dados também podem ser reutilizados dentro de uma estrutura conceitual diferente, por amostragem ou os reestruturando para a combinação com outros conjuntos e usos em produções diversas.

O modelo de ciclo de vida de dados DDI, de forma simples e resumida, destaca o processo de investigação, visivelmente nas três primeiras etapas. As práticas para curadoria e para a RDM estão representadas nas demais fases do modelo, podem ser percebidas nas etapas do arquivamento e reaproveitamento, ambas dedicadas à preocupação com o tratamento e a partilha dos dados. Este modelo também tem sido base para elaboração de outras estruturas de ciclos de vida, ou mesmo usado na íntegra por variadas instituições e

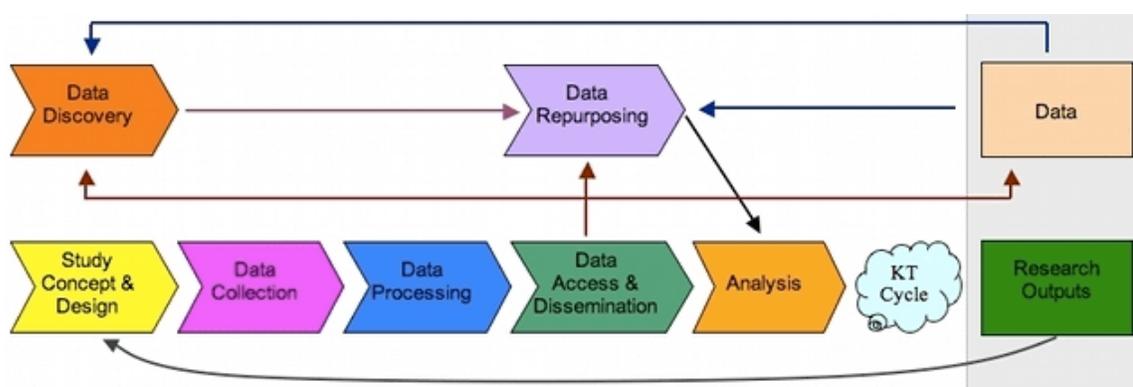
universidades⁴⁰. Isto parece ocorrer devido à adoção do modelo por diversos membros da aliança DDI, mas também por ser uma representação que traz na sua constituição aspectos comuns da investigação, o que parece tornar mais fácil a visualização e associação por seus usuários. É na simplicidade da ação sequencial e nos fluxos alternativos que a estrutura da DDI pode contribuir com o modelo desta investigação, pois a concepção mostra-se relevante, tendo em vista a adoção de sua base por outros modelos, incluindo a *e-Science and the lifecycle of research*, uma influente estrutura de uso em nível teórico e prático, amplamente citada na literatura (Cox & Tam, 2018) e mostrado a seguir.

4.2.3. E-SCIENCE AND THE LIFECYCLE OF RESEARCH

O modelo do ciclo de vida intitulado de “criação do conhecimento na investigação” (Figura 16), foi divulgado no trabalho de Humphrey (2006) – *e-Science and the lifecycle of research*. Apesar do título, têm os dados como foco principal. Preparado para a *Association of Research Libraries* do Canadá, a estrutura do modelo assemelha-se ao do modelo DDI, em termos gráficos e na descrição das etapas. No entanto, a principal contribuição deste modelo para a tese, encontra-se nas peculiaridades em relação aos seus fundamentos.

Figura 16

Modelo do ciclo de vida e-Science and the lifecycle of research



Nota. Elaborado por Humphrey (2006).

⁴⁰ MIT - Massachusetts Institute of Technology - https://ocw.mit.edu/resources/res-str-002-data-management-spring-2016/workshop-materials/MITRES_STR002S16_IntroDM.pdf.

UCL - University College London - https://www.closer.ac.uk/wp-content/uploads/3.1-DDI-C_and_DDI-L.pdf.

UA - University of Alberta (Humphrey, 2008).

UVa - University of Virginia - <https://guides.lib.virginia.edu/c.php?g=515290&p=3522215>.

O modelo apresentado é constituído por dez fases: 1) *Study Concept & Design*; 2) *Data collection*; 3) *Data Processing*; 4) *Data Access & Dissemination*; 5) *Analysis*; 6) *KT Cycle*; 7) *Research Outputs*; 8) *Data Discovery*; 9) *Data Repurposing*; e, 10) *Data*. Em sua maioria, estas fases já estão contemplados e descritas no modelo da DDI, as distinções entre eles, que podem ser percebidas de forma discreta, estão principalmente nas nomenclaturas e no acréscimo das etapas *KT Cycle*, *Research Outputs* e *Data*.

Para conhecer os fundamentos deste modelo, é necessário uma breve descrição das práticas realizadas nas etapas, tal como segue: o fluxo dos dados e da investigação inicia-se com o projeto (*Study Concept & Design*), seguido por uma coleta dos dados para uso na investigação (*Data collection*) através do processamento dos dados (*Data Processing*), de modo a produzir dados possíveis de serem acessados e disseminados (*Data Access & Dissemination*). Acima da fase de acesso e disseminação, existe um *loop* de reaproveitamento (*Data Repurposing*), onde é iniciado um processo circular para que os dados disseminados sejam descobertos (*Data Discovery*) e estejam disponíveis para serem analisados e reutilizados. Neste sentido, os dados disseminados por meio do reaproveitamento ou de forma direta pela investigação, estarão disponíveis para serem analisados também por terceiros (*Analysis*). Um novo ciclo é sugerido a partir dos resultados da investigação (*Research Outputs*), estimulando um reinício do ciclo de vida.

Humphrey (2006) observa também que a linha superior do modelo está associada ao paradigma da *e-Science*, pois envolve as relações entre os dados (*Data*), descoberta (*Discovery*) e reaproveitamento destes dados, em um vínculo fundamental com os resultados da investigação. O autor exemplifica que o fluxo para o reaproveitamento pode representar a mineração de dados e de texto, que são atividades reconhecidas da *e-Science*. O modelo também propõe uma fase de transferência de conhecimento (*knowledge transfer - KT*), elaborada para proporcionar fácil acesso aos dados e às informações por trás das descobertas da investigação, pois assim, segundo o autor, se enriquece a experiência de transferência de conhecimento para futuros investigadores e leitores da investigação.

O modelo do ciclo de vida *e-Science* tem como fundamento proporcionar um melhor entendimento das relações entre as etapas da investigação, e ainda aumentar a consciência das perdas potenciais de informação entre estas fases. Nesse sentido, o autor

destaca as lacunas entre as etapas como os locais mais vulneráveis à perda de informações, definindo-as como pontos de transição. Ou seja, os pontos de transição são atividades na investigação que fluem ao longo do ciclo de vida e ocorrem na medida que se passa de uma etapa para a etapa seguinte, fazendo das junções entre as fases elementos cruciais de atenção para o modelo. A título de exemplo, o autor explica que na transição do estágio de *Data Collection* para o *Data Processing*, a passagem de questionários realizados na investigação, pode não incluir as fontes das perguntas ou o método de seleção para a amostra coletada, ou ainda, detalhes sobre como administrar o questionário e informações essenciais para o uso em etapas posteriores, sendo assim possível ocasionar potenciais perdas de informação.

Os pontos de transição também podem incluir atividades como administração de projetos, licitações, gerenciamento de dados, criação de conhecimento, julgamentos éticos, supervisão de propriedade intelectual e gerenciamento de tecnologia. Isto implica que a transferência das informações entre estas fases, envolve mudanças de responsabilidade de uma pessoa para outra ou de uma organização para outra. Assim, partes interessadas como agências de fomento, universidades, investigadores, produtores de dados, editoras, bibliotecas e arquivos de dados, também formam o conjunto de atores considerados nestes pontos de transição observados no ciclo de vida da *e-Science*.

Embora semelhante ao modelo DDI, este ciclo de vida apresenta como diferencial uma ampliação do seu escopo, a partir do ponto de vista dos dados gerados, enfatizando as perdas potenciais de informação no processo. Isto pode ser percebido na ideia da transferência de conhecimento da etapa nomeada como *KT Cycle* e nos pontos de transição, indicando a atenção que deve ser considerada entre cada fase do ciclo de vida. Outro destaque encontra-se na imagem que aponta para a investigação prolongada em novos ciclos, indicando que o “arquivamento” dos dados e os resultados da investigação, não são o fim do processo, mas participantes cíclicos da investigação, como afirma o próprio autor juntamente com Jacobs (Jacobs & Humphrey, 2004):

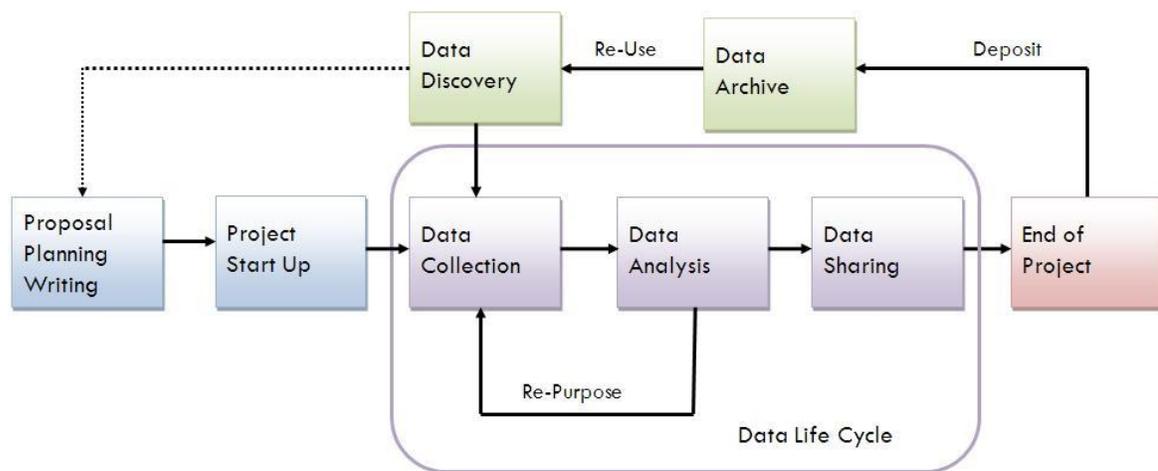
Data archiving is a process, not an end state where data is simply turned over to a repository at the conclusion of a study. Rather, data archiving should begin early in a project and incorporate a schedule for depositing products over the course of a project's life cycle and the creation and preservation of accurate metadata, ensuring the usability of the research data itself. Such practices would incorporate archiving as part of the research method (p.28).

A possível contribuição do modelo da *e-Science* para esta tese pode ser vista em dois sentidos. O primeiro sentido está na necessidade da transferência de conhecimento e o segundo na idealização de pontos de transição, como marco de atenção para vulnerabilidades e possíveis perdas de dados. A inclusão de tais sentidos pode servir para conjugar as práticas essenciais para a RDM com as responsabilidades dos atores envolvidos, de modo a possibilitar a especificação e o desenvolvimento dos relacionamentos entre esses elementos. Estes dois sentidos, embora importantes para as práticas para a RDM e curadoria, são difíceis de perceber no tocante ao momento em que podem ser realizados no processo, mas também quanto à sua organização no ambiente da investigação.

4.2.4. UVa - VIRGINIA UNIVERSITY

Figura 17

Modelo do ciclo de vida da UVa - University of Virginia



Nota. Elaborado por University of Virginia Library (2022).

O modelo do ciclo de vida da Universidade da Virgínia – UVa (Figura 17), foi desenvolvido por um grupo de consulta de gerenciamento de dados da biblioteca daquela universidade. A biblioteca da UVa em seu escopo de trabalho fornece informações aos investigadores sobre a RDM e todo o ciclo de vida, com recursos para criar, gerenciar, armazenar e citar os dados de investigação, bem como partilhá-los e preservá-los para o futuro. O objetivo é tornar a RDM mais fácil, com informações e treinamento sobre as

melhores práticas, planos de gestão, armazenamento, segurança e repositórios de dados (University of Virginia Library, 2022).

O diagrama da UVa é constituído por oito etapas. A interpretação de cada etapa do modelo, assim como as práticas relacionadas, podem ser percebidas sob o referenciamento elaborado por Carlson (2014), da seguinte forma:

1) ***Proposal Planning & Writing*** (Planejamento e redação de propostas): esta etapa está ligada à descoberta dos dados através da linha pontilhada da figura e tem por objetivo: realizar uma revisão dos conjuntos de dados existentes; determinar se o projeto produzirá um novo conjunto de dados ou combinará os dados existentes; identificar os usuários em potencial; determinar os custos relacionados ao arquivamento dos dados; bem como reconhecer os desafios para o arquivamento; consentimento; e, confidencialidade; e

2) ***Project Start Up*** (Arranque do projeto): uma vez que o projeto é definido, este estágio pode incluir atividades como a compra de equipamentos, contratação de pessoal e um planeamento da atividade. De modo prático, estas ações incluem criar um plano de gestão de dados, definir a forma e o conteúdo da documentação, realizar pré-teste e testes nos materiais e métodos para a investigação.

As próximas três fases estão nomeadas de forma específica como: o ciclo de vida dos dados, inicia-se com a etapa de coleta, passa pela análise e termina na partilha dos dados.

3) ***Data Collection*** (Colecção de dados): durante esta fase são gerados, coletados ou adquiridos os dados necessários para conduzir a investigação, com ações relacionadas para a organização dos arquivos, armazenamento e *backups*, como também controlar a qualidade da coleta, acesso e segurança dos dados, seguindo as melhores práticas;

4) ***Data Analysis*** (Análise de dados): concluída a coleta, os dados são utilizados durante esta fase de análise para gerar as informações necessárias ao suporte da investigação. O gerenciamento de versões e a manipulação de arquivos, assim como a análise de documentos, também são práticas aqui atribuídas. A seta que indica o reaproveitamento sugere que, se forem descobertos erros, os dados precisarão ser coletados novamente;

5) **Data Sharing** (Partilha de dados): após a análise, os dados são compartilhados em modos diferentes, seja indiretamente por meio de apresentações, publicações e outros produtos da investigação, ou diretamente por meio de acordos individuais. Esta fase conta com práticas para determinar formatos, contatar arquivos, limpar e documentar os dados; e

6) **End of Project** (Fim do projeto): aqui a investigação é concluída e as práticas são executadas para o encerramento do projeto. Tais práticas podem ser realizadas com a geração de artigos e relatórios finais para as agências de fomento, comunicação ou partes interessadas, e o depósito em arquivos de dados. Assim, os dados tornam-se disponíveis, mantêm seu valor além da vida do projeto, podem ser descobertos e reaproveitados. Nesse sentido geram-se novas ideias para serem reutilizadas em projetos de investigação, completando e reiniciando o ciclo de vida retratado neste modelo.

O modelo da UVa segue a mesma tendência visual do modelo DDI, mas caracteriza-se por ser desenvolvido com a visão de uma biblioteca universitária. Este modelo se destaca por deixar muito bem identificadas graficamente as ações do processo de investigação na tríade coleta-análise-partilha, do mesmo modo o reinício do ciclo de vida com o reaproveitamento dos dados, que, neste caso, refere-se aos dados dentro do processo da própria investigação. O modelo também ratifica de forma nítida que o ciclo de vida tem como objetivo final a descoberta e a reutilização dos dados, a resultar na possibilidade de geração de novas ideias e reinício do fluxo. As práticas de curadoria e para a RDM, incluindo a preservação e a partilha dos dados, estão caracterizadas de modo a ser facilmente identificadas pela nomenclatura, mas também sob as explicações disponibilizadas pela biblioteca, através de um guia *online* para a RDM que oferece recursos para seguir o fluxo do ciclo de vida, ainda acrescentando outras práticas para a gestão dos dados.

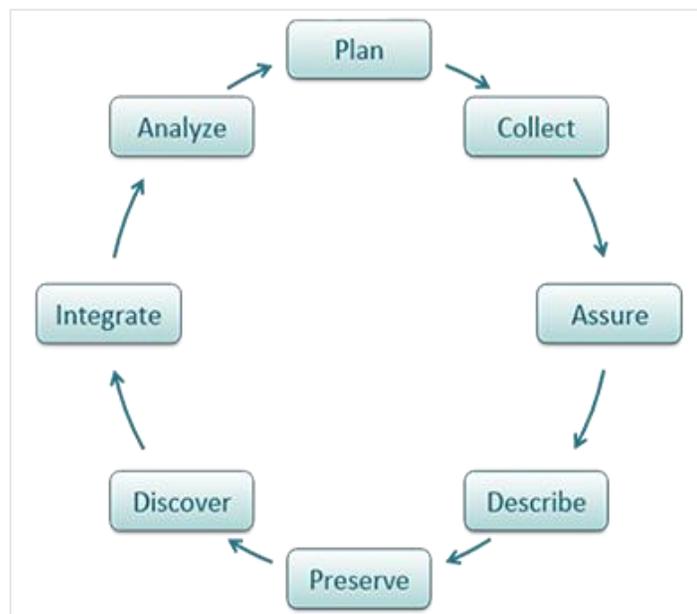
O modelo desenvolvido pela biblioteca da UVa, pode ser importante e distintivo para esse estudo em pelo menos duas perspectivas: a clareza com que demonstra a separação e a identificação das etapas no que envolve o processo de investigação, a curadoria e os outros componentes que estão no ambiente da RDM, além da disponibilização dos recursos que devem acompanhar os modelos para facilitar sua identificação e uso. Pondera-se assim que estas duas perspectivas generalistas podem ajudar no acolhimento da RDM, sob a diversidade disciplinar que caracteriza o ambiente da universidade.

4.2.5. DATAONE - DATA OBSERVATION NETWORK FOR EARTH

O DataONE (*Data Observation Network for Earth*) é um programa com sede nos Estados Unidos, conduzido por uma comunidade internacional de investigadores, gerentes de informação, bibliotecários, autores de dados, usuários e outras partes interessadas. O objetivo da instituição é fornecer acesso e promover um guia de melhores práticas em gerenciamento de dados, por meio de recursos e materiais educacionais responsivos, objetivados para melhor compreender e conservar a vida no planeta e o meio ambiente. De modo prático, o programa disponibiliza vários repositórios para a partilha dos dados, por meio de uma ciberinfraestrutura distribuída que visa atender as necessidades da ciência e da sociedade para o acesso persistente, robusto, seguro e aberto. Com o apoio da *National Science Foundation* (NSF), este modelo do ciclo de vida (Figura 18) foi desenvolvido em colaboração com a comunidade DataONE (Data Observation Network for Earth [DataOne], 2021, 2021a, 2021b; Strasser et al., 2012).

Figura 18

Modelo do ciclo de vida DataONE



Nota. Elaborado por DataONE (2021).

A estrutura fornece uma visão geral dos estágios envolvidos na gestão e preservação destinados ao uso e reuso dos dados, servindo como uma estrutura básica para

orientar o desenvolvimento e a utilização de recursos, bem como serviços e materiais. As etapas do modelo podem ser expandidas, possibilitando inclusive a inserção de novos fluxos. Nesse sentido, o *DataONE Investigator Toolkit* é um recurso que permite aos investigadores acessarem as ferramentas ali contidas para que o apoio seja fornecido em todas as fases do ciclo de vida (DataOne, 2021, 2021a, 2021b; Strasser et al., 2012).

Neste modelo circular, as etapas podem ocorrer de uma vez ou em várias vezes, em sequências diferentes, com algumas fases ocorrendo ao mesmo tempo e algumas sendo repetidas com maior frequência do que outras. Isto não exige que os projetos utilizem obrigatoriamente todos os componentes, permitindo o uso de apenas parte do ciclo e não necessariamente o caminho linear apresentado em formato circular (Strasser et al., 2012). São variações de acordo com as práticas, domínios e necessidades das comunidades de uso.

O modelo apresentado pela DataOne mostra uma visão abrangente dos estágios que consistem em oito etapas, a saber:

1) **Plan** (Planejar): esta etapa traduz-se na elaboração de um plano de gestão dos dados para o apoio das políticas e da sustentabilidade. É a descrição de como os dados serão compilados, gerenciados e disponibilizados ao longo da vida útil;

2) **Collect** (Coletar): consiste na fase em que os dados primários são coletados, considerando os métodos e a documentação para a sua recolha e as observações necessárias sobre o formato digital. Estas práticas visam garantir a usabilidade posterior e uma melhor implementação da gestão dos dados;

3) **Assure** (Assegurar): são os critérios e procedimentos que asseguram a qualidade e a validação dos dados, identificando possíveis erros durante a coleta, assim como o emprego de técnicas para solucioná-los, por meio de verificações e inspeções. Problemas com os dados também podem ser identificados durante a fase da análise para a interpretação;

4) **Describe** (Descrever): contempla a descrição minuciosa dos dados por meio da adoção de padrões e ferramentas de metadados para a partilha e reutilização dos dados. Nesta etapa as ferramentas e *softwares* podem estar disponíveis para ajudar a descrever os dados de forma precisa, usando os padrões apropriados;

5) **Preserve** (Preservar): através do envio dos dados para um arquivo pertinente, do tipo *data centers* ou repositórios, realiza-se a preservação dos dados em curto e em longo

prazo. Em curto prazo para minimizar as perdas potenciais, em longo prazo para que as partes interessadas do projeto possam acessar, interpretar e usar os dados no futuro;

6) **Discover** (Descobrir): configura-se na identificação de outros conjuntos de dados para complementar e agregar valor à investigação. São os dados localizados e obtidos juntamente com os metadados. É necessário uma partilha adequada para garantir que os conjuntos de dados obtidos sejam reproduzíveis, vistos ou acessados por qualquer pessoa;

7) **Integrate** (Integrar): é a combinação dos dados de fontes distintas, para formar um conjunto homogêneo para serem analisados. Para que haja a possibilidade de integração, é necessário que os dados sejam compatíveis e comparáveis; e

8) **Analyze** (Analisar): consiste na exploração, verificação e visualização dos dados, de modo a identificar padrões, testar hipóteses e ilustrar descobertas para tornar esses dados compatíveis, comparáveis e reproduzíveis, consequentemente reutilizáveis.

O modelo DataOne é uma representação utilizada para documentar o estado real da RDM (Weber & Kranzlmülle, 2019). Um aspecto a ser evidenciado é o fato de que nem todas as etapas do ciclo de vida seguem um caminho sequencial, de acordo com as orientações do modelo, isto depende do tipo de uso a ser realizado. Assim, o modelo pode combinar vários e diferentes fluxos de trabalhos em uma única expressão circular. Ball (2012) exemplifica assim: uma investigação que envolve apenas novos dados prosseguirá através das etapas 1 a 5 (planejar, coletar, assegurar, descrever e preservar) e em seguida, pode passar direto para a etapa 8 (analisar). Outro exemplo, é que uma síntese ou meta-análise pode seguir um caminho que consiste nas etapas 6, 3, 7 e 8 (descobrir, assegurar, integrar e analisar), podendo haver em ambos os casos, alguma sobreposição entre as etapas.

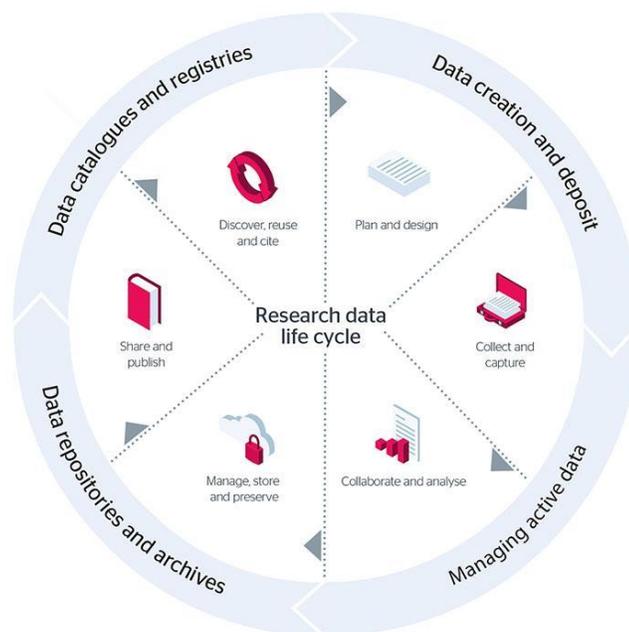
Mais um aspecto a ser observado no modelo é a maior ênfase nos dados e o modo de gerenciá-los, e a menor ênfase nos aspectos do processo de investigação, de modo que o modelo utiliza os nomes semelhantes das etapas da investigação, trazendo à primeira vista uma certa confusão sobre o conteúdo representado na fase. Este modelo pode servir para apontar a importância da curadoria na atividade da RDM, onde as etapas não precisam ser necessariamente sequenciais, podendo inclusive serem definidas e expandidas em fluxos auxiliares. No entanto, o modelo não deve ser realizado sem a perda da atenção para o delineamento da investigação e o seu processo consequente.

4.2.6. JISC - JOINT INFORMATION SYSTEMS COMMITTEE

O Jisc (*Joint Information Systems Committee*) é uma organização sem fins lucrativos, que fornece aos setores da educação superior e da investigação no Reino Unido assistência em soluções e serviços de infraestrutura digital partilhada. O modelo do ciclo de vida da Jisc (Figura 19) tem a sua base na curadoria digital através das fases do processo de investigação (Beagrie, 2004). A estrutura é apresentada e explicada juntamente com um Kit de ferramentas desenvolvido para a RDM, de acordo com os estágios do modelo. O diagrama é composto por seis etapas sequenciais, envolvidas por quatro práticas para a RDM. Conforme o detalhamento que segue (Jisc, 2021):

Figura 19

Modelo do ciclo de vida do Jisc



Nota. Elaborado por Jisc (2021).

1) **Plan and design** (Planear e projetar) – é necessário considerar os dados no início do projeto e antes de iniciar o trabalho prático, também as políticas, serviços e requisitos relevantes no contexto organizacional, ajudando a minimizar os custos e a garantir que os dados possam ser organizados, armazenados e preservados com uso dos princípios FAIR (*Findable, Accessible, Interoperable and Reusable*) e partilhados no momento adequado com as pessoas certas e permissões apropriadas;

2) ***Collect and capture*** (Coletar e capturar): após elaborado o planejamento e iniciado o projeto, usa-se um sistema adequado para a coleta e armazenamento dos dados, junto com as soluções de *backup*, documentação e metadados apropriados, de modo a permitir que outras pessoas encontrem, contextualizem, reutilizem e citem o trabalho. Os formatos de arquivo, nomenclatura e estruturas de pastas apropriadas devem ser projetados para localizar informações e compatibilidades futuras na reutilização dos dados;

3) ***Collaborate and analyse*** (Colaborar e analisar): com os dados coletados, os métodos de análise dos dados e as convenções utilizadas devem ser documentadas de modo a permitir que os colaboradores reutilizem os dados. Alerta-se que a partilha requer atribuições de funções e responsabilidades, uso de plataformas de colaboração e comunicação de forma pertinente, de modo a permitir o controle e o armazenamento dos dados com os cuidados necessários para uso das partes interessadas;

4) ***Manage, store and preserve*** (Gerenciar, armazenar e preservar): esta etapa trata de como proteger os dados durante o projeto, inclui considerações legais e técnicas sobre a segurança, preservação e a proteção dos dados. É importante considerar que o local de armazenamento afeta o gerenciamento, bem como os formatos dos arquivos e a preservação. Assim, é preciso atentar rotineiramente aos aspectos desta etapa;

5) ***Share and public*** (Partilhar e publicar): os dados podem ser partilhados informalmente durante o projeto ou formalmente no final do projeto, via repositórios ou material complementar às publicações. A execução desta etapa significa possibilitar a descoberta e o uso dos dados com mais facilidade para que possam ser citados através de identificadores persistentes, replicados e validados. Neste estágio as licenças devem ser consideradas, do mesmo modo os termos, condições ou obrigações e impedimentos legais para a partilha, isto envolve normas disciplinares, infraestruturas técnica e treinamento; e

6) ***Discover, reuse and cite*** (Descobrir, reutilizar e citar): na RDM estas práticas podem ser vistas em duas perspectivas, a gerência do processo ou o consumo dos dados. A perspectiva desta etapa concentra-se na função consumo dos dados, ou seja, aqui os dados são encontrados em repositórios, ou seguindo os links em um artigo, com o objetivo da reutilização. Os termos e condições para a reutilização dos dados são verificados nesta etapa.

Além do significado das etapas descritas anteriormente, o Jisc também detalha o conteúdo destas fases. Nesse sentido, a tabela abaixo apresenta as etapas e os conteúdos descritos de forma resumida, que constituem o ciclo de vida visualizados no modelo Jisc.

Tabela16

Descrição do conteúdo das ações que constituem o modelo Jisc

Etapas	Conteúdo	Descrição
1 - Planejar e projetar	Plano de gestão dos dados	Descrevem como os dados são gerenciados e documentados ao longo de um projeto de investigação. Ele também discute o depósito para preservação em longo prazo.
	Conformidade política	É a verificação do cumprimento das políticas relevantes dos financiadores e das organizações internas e externas.
	Princípios FAIR	Os princípios FAIR garantem que os resultados da investigação sejam localizáveis, acessíveis, interoperáveis e reutilizáveis. Na prática, permitem uma maior visibilidade e potencial para a reutilização dos dados.
	Custeio	É o investimento de tempo e dinheiro que dever ser realizado para custear a RDM de forma adequada, com um plano de gerenciamento de dados.
2 - Coletar e capturar	Armazenamento de dados ativos e <i>backup</i>	Antes de iniciar o trabalho prático, soluções de armazenamento e <i>backup</i> devem ser selecionadas. Isto significa que os dados estão seguros durante um projeto de investigação, inclusive no caso de problemas imprevistos.
	Catálogo dos dados - questões de metadados	Uma descrição dos dados da investigação através de metadados ajuda a interpretar e a reutilizar as descobertas, fornecendo informações e percepções sobre o trabalho e a metodologia utilizada.
	Gerenciamento e formatos de arquivos	Antes de começar a coletar dados, uma hierarquia de pastas e convenções de nomenclatura devem ser escolhidas e documentadas. Além disso, os formatos dos arquivos devem ser considerados para permitir a reutilização em longo prazo.
3 - Colaborar e analisar	Investigação colaborativa	Significa partilhar recursos, equipamentos, dados, entre outros. As plataformas colaborativas e uma ampla variedade de serviços de partilha de arquivos devem estar disponíveis para ajudar os investigadores na partilha.
4 - Gerenciar, armazenar e preservar	Preservação	A preservação de dados visa manter o conteúdo desses dados e o seu contexto (metadados) seguros em uma possível reutilização futura, para facilitar a preservação, deve-se considerar o planejamento.
	Segurança	As pessoas autorizadas devem ter acesso aos dados da investigação. Portanto, os dados devem ser protegidos por meio de ferramentas adequadas, que inclui medidas para proteção nos piores cenários.
	Proteção de dados	A proteção de dados trata de resguardar todas as informações que precisam ser custodiadas. Existem regras específicas a serem seguidas e multas graves em caso de incumprimento.
	Proteção de informações pessoais	As informações de identificação pessoal precisam ser tratadas com cuidado. Embora indispensável em muitas áreas, o uso indevido pode resultar em multas e questões éticas.

	Software	O <i>software</i> é frequentemente criado para analisar e manipular os dados. Assim, os <i>softwares</i> também precisa ser gerenciados, armazenados e preservados para permitir a reutilização futura.
5 - Compartilhar e publicar	Propriedade intelectual e direitos autorais	Os criadores dos dados recebem determinados direitos de propriedade intelectual. Portanto, para permitir que outros reutilizem o trabalho, deve-se escolher uma licença apropriada.
	Licenciamento	O licenciamento permite que outros reutilizem os dados da investigação. A licença declara o que os utilizadores podem fazer com o trabalho, incluindo a citação e a partilha.
	Identificadores	Os identificadores permitem que os investigadores citem e sejam citados. Como links persistentes para o conteúdo que garanta ao objeto ser detectável sempre, mesmo se sua localização mudar.
	Vinculação dos dados a uma publicação	As publicações de investigação são mais úteis quando apoiadas pelos dados subjacentes. Existem caminhos para partilhar dados com publicações, incluindo repositórios <i>online</i> , periódicos regulares e periódicos especializados em dados.
6 - Descobrir, reutilizar e citar	Citação de dados e métricas	Os dados devem ser citados como publicações acadêmicas, esta prática rende citações adicionais, além de reconhecer os esforços dos investigadores e permitir maior transparência.

Nota. Adaptado do Jisc, (2021).

O modelo do ciclo de vida dos dados Jisc tem sua ênfase na RDM e na curadoria digital, pois cita e explica diversos elementos característicos da RDM e detalha as práticas de curadoria em várias etapas do seu ciclo. As etapas do modelo são envolvidas por uma camada exterior com quatro práticas ou “preocupações” para a RDM: 1) a criação e o depósito dos dados (*Data creation and deposit*); 2) o gerenciamento ativo dos dados (*Managing active data*); 3) os repositório e arquivos dos dados (*Data repositories and archives*); e, 4) a catalogação e registro dos dados (*Data catalogues and registries*).

A camada exterior com as quatro práticas permite inferir de forma mais clara sobre a necessidade de atores diferentes do investigador para a execução das ações propostas, pois não são ações principais do processo de investigação⁴¹. Além da representação gráfica em torno da RDM e do processo de investigação, o modelo apresenta um material para que as práticas para a RDM sejam implementadas, pois juntamente com o modelo é fornecida uma lista de checagem que permite conhecer as práticas e destacar as ferramentas para execução em cada fase do ciclo, de modo a orientar como a RDM e a curadoria devem acontecer.

⁴¹ Obviamente que o investigador é o ator central das práticas, mas nem todas as ações são exclusivas do investigador.

Pelas especificidades do modelo, é possível perceber que princípios e práticas da ciência aberta em relação a partilha permeiam todas as suas fases, também a necessidade de designar atores para a realização das práticas que excedem o trabalho do investigador no processo de investigação. A organização das práticas, acrescidas da designação de atores responsáveis, além da própria ênfase na RDM e na curadoria, podem ser consideradas norteadoras e fundamentais para a composição do modelo resultante deste trabalho.

4.2.7. UK DATA ARCHIVE – UNIVERSITY OF ESSEX

A Universidade de Essex abriga o *UK Data Archive (United Kingdom Data Archive - UKDA)*, um centro de especialização em dados, importante provedor de serviços direcionados às práticas da RDM. A instituição, em parceria com a *UK Data Service*⁴², dispõe de recursos de investigação e infraestrutura técnica para curar, fornecer e gerenciar o acesso seguro e em longo prazo aos dados econômicos, sociais e populacionais do Reino Unido, fornecendo também treinamento e suporte para o uso de repositórios de dados e outros recursos para a RDM. O centro abrange parcerias com as universidades locais e internacionais, departamentos governamentais e centros de arquivos de dados⁴³. Estas parcerias têm por objetivo atender às necessidades digitais, atuais e futuras, dos investigadores, usuários e proprietários de dados, mas também partilhar o conhecimento, experiências e treinamento na área (UK Data Archive, 2021; UK Data Service, 2021a).

O modelo disponibilizado pelo UKDA é direcionado às atividades de suporte, orientação e serviços da RDM desenvolvidos pela instituição. Essas atividades estão organizadas ao longo das fases do ciclo de vida. O escopo relacionado aos dados de investigação apresenta-se em uma estrutura que propõe ajudar os investigadores a associar

⁴² A UK Data Service é nomeada como uma infraestrutura de investigação. Em 2020, hospedou e forneceu acesso online aos primeiros conjuntos de dados da COVID-19, permitindo que investigadores e desenvolvedores de políticas no Reino Unido, gerassem respostas sobre os impactos econômicos e sociais da pandemia (UK Data Service, 2021)

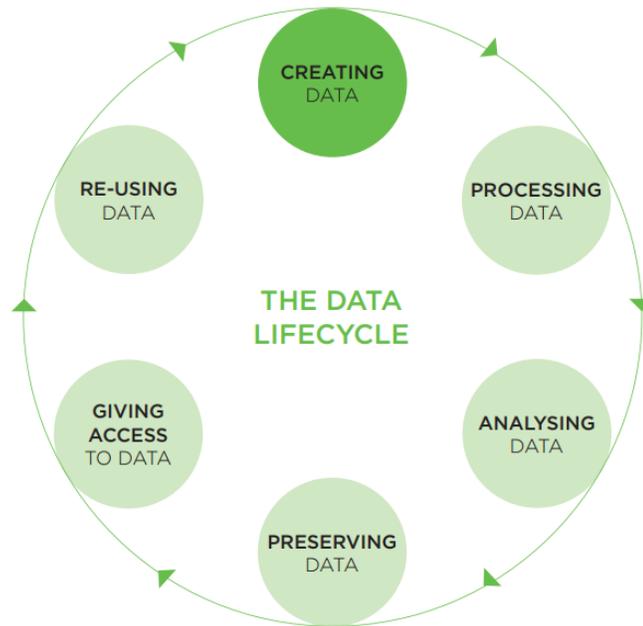
⁴³ Universidades de Manchester, Southampton, Londres, Edimburgo, bem como Jisc, CESSDA - Consortium of European Social Science Data Archives, IDAN - International Data Access Network, FAIRs FAIR - Fostering Fair Data Practices in Europe, IASSIST –International Association for Social Science Information Services & Technology, ICPSR – The Inter-university Consortium for Political and Social Research, e, treinamentos na Europa, EUA, China, Índia, Taiwan e Quirguistão (UK Data Archive, 2021).

as práticas recomendadas para a RDM com o processo da investigação, através dos recursos textuais e operacionais que acompanham o modelo do UKDA (Corti et al, 2014).

São seis as fases que representam e identificam as práticas a serem desenvolvidas a partir do modelo do *UK Data Archive* (Figura 20), conforme a explicação que segue.

Figura 20

Modelo do ciclo de vida do UK Data Archive.



Nota. Elaborado por Corti et al (2014).

Para entender o modelo do UKDA, Van den Eynden (2013) exemplifica seu uso através de duas investigações. A primeira foi realizada na área de saúde sobre as consequências sociais da epidemia de febre aftosa ao norte de Cumbria - Reino Unido (Mort, 2006). O segundo exemplo trata sobre o projeto de arquivo métrico de arqueologia de Wessex, que reuniu dados de ossos de animais de sítios arqueológicos na Inglaterra (Grimm, 2008). O segundo exemplo é utilizado para destacar que a associação das fases da investigação arqueológica são realizadas em cinco etapas no modelo, de modo a sugerir a existência de flexibilidade e adaptabilidade no modo a ser seguido. Já no primeiro exemplo, a autora relaciona de maneira simples e direta algumas práticas do processo de investigação, associados aos estágios sequenciais do modelo. A idealização das etapas, tendo por base o primeiro exemplo, é explicado em ações, na seguinte narrativa da autora referenciada:

1) **Creating Data** (Criação dos dados): o projeto de investigação foi elaborado com o uso de diários e gravação para entrevistas, juntamente com o consentimento para participação e uso dos dados primários;

2) **Processing Data** (Processamento): posteriormente à realização das entrevistas, os diários foram transcritos;

3) **Analysing Data** (Análise): o arquivamento dos dados foram discutidos e o consentimento institucional para arquivar as transcrições e gravações obtidos;

4) **Preserving Data** (Preservação): as transcrições e as gravações foram arquivadas e o registro do catálogo com o guia de usuário criados;

5) **Giving Access to Data** (Acesso): forma disponibilizados as transcrições e o guia do usuário, no caso usou-se o mesmo do UKDA; e

6) **Re-Using Data** (Reutilização): os dados foram reutilizados em outros estudos.

No exemplo da narrativa, os dados foram reutilizados em estudos sobre a avaliação de fontes de conhecimento no controle de doenças animais. Uma visão geral e mais detalhada das etapas, encontra-se na lista de práticas a serem realizadas a partir das fases do modelo, identificadas na tabela a seguir:

Tabela 17

Etapas e práticas do modelo do UKDA

Fases do modelo do UK Data Archive	Práticas a serem realizadas a partir das etapas
1. Criar os dados	<ul style="list-style-type: none">• conceber a investigação;• planejar o gerenciamento de dados (formatos, armazenamento etc.);• planejar o consentimento para a partilha;• localizar os dados existentes;• coletar os dados (experimentar, observar, medir, simular); e• capturar e criar metadados.
2. Processar os dados	<ul style="list-style-type: none">• inserir os dados, digitalizar, transcrever e traduzir;• verificar, validar e limpar os dados;• anonimizar os dados quando necessário;• descrever os dados; e• gerenciar e armazenar os dados.
3. Analisar os dados	<ul style="list-style-type: none">• interpretar os dados;• derivar os dados;• produzir resultados de investigação;• publicações do autor; e• preparar os dados para a preservação.

4. Preservar os dados	<ul style="list-style-type: none"> • migrar os dados para o melhor formato; • migrar os dados para um meio adequado; • fazer <i>backup</i> e armazenar os dados; • criar metadados e documentação; e • arquivar os dados.
5. Dar acesso os dados	<ul style="list-style-type: none"> • distribuir os dados; • partilhar os dados; • controlar o acesso; • estabelecer direitos autorais; e • promover os dados.
6. Reutilizar os dados	<ul style="list-style-type: none"> • acompanhar a nova investigação; • realizar análises de investigação; • examinar as descobertas; e • ensinar e aprender.

Nota. Adaptado de Corti et al (2011).

As ações referenciadas e listadas na tabela 17 possuem os elementos necessários para perceber, de modo básico, como as práticas são direcionadas para a realização da curadoria e para a RDM. Cabe observar que na versão e organização do material da instituição, os assuntos relacionados à RDM estão dispostos sem forte ênfase no modelo do ciclo de vida (UK Data Service, 2021b; Weber & Kranzlmüller, 2019)

Segundo Cox e Tam (2018), o modelo é centrado em dados e no processo da investigação, porém não mostra o contexto mais amplo, ou seja, de como um projeto é concebido ou planejado, o que pode dificultar o vínculo do investigador com o processo. Ao observar os elementos e o encadeamento do modelo, os autores percebem um baixo nível de abstração, praticamente fechado, unidirecional e serial, consistindo em fases compostas por práticas que ocorrem em cada estágio. Embora o modelo seja circular, pode haver uma interrupção entre etapas em termos de reinício do ciclo de vida, pois, ainda segundo os autores, há uma diferença descontínua entre tornar os próprios dados acessíveis (etapa 5), o que indica o fim da atividade do projeto e, em seguida, procurar outros dados para a reutilização (etapa 6), o que pelo modelo implicaria no início de um novo projeto.

No modelo do UKDA um ponto relevante a ser considerado é a ênfase nos dados e a sua associação direta com o processo da investigação. Porém, diferente dos outros modelos deste tipo aqui estudados, o diagrama é circular e resumido em seis etapas – criação, processamento, análise, preservação, acesso e reutilização. O contexto e os recursos para a RDM naquela instituição, podem ser úteis para o apoio e orientação na realização das práticas para a RDM. Porém faltam, a partir do modelo, etapas e práticas que possam

direcionar e explorar a RDM e a curadoria. A estrutura e o contexto deste modelo também podem ser úteis na relação enfática com o processo da investigação. Por outro lado, as observações que não são totalmente positivas, contribuem também para que a projeção do modelo aqui proposto, de modo que, quando consideradas pertinentes, sejam evitadas.

4.2.8. CENS - CENTER FOR EMBEDDED NETWORKED SENSING

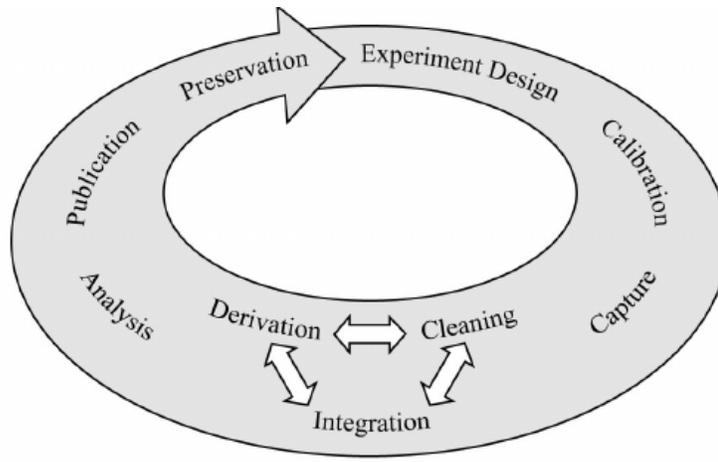
O CENS (*Center for Embedded Networked Sensing*) foi estabelecido em 2002 para desenvolvimento de sistemas de detecção sem fio e suporte multidisciplinar a professores, alunos e funcionários em diversas áreas do conhecimento de universidades localizadas nos Estados Unidos. A elaboração do modelo do CENS foi realizada a partir de cinco projetos das ciências ambientais, tendo por direcionamento questões sobre a natureza dos dados, como estes são produzidos e gerenciados. Mais especificamente, o modelo é resultado de entrevistas e observações de campo na área da ecologia ambiental com tecnologia de detecção em rede. Assim, o modelo do CENS envolve dados criados nas ciências, nas tecnologias e na combinação das tecnologias para as ciências (Wallis et al., 2008).

Segundo Wallis et al. (2008), a heterogeneidade do contexto fez com que fosse inicialmente idealizada a criação de um modelo unificado para todas as investigações, mas tal ideia foi posteriormente alterada. Nesse sentido os autores justificam assim: “There was a clear distinction between what happened to the data created as a part of technology research and those data created when science and technology researchers worked together, and these were both different from the data created by scientific research” (p.3). Para adequar as especificidades, o modelo do CENS foi concebido com nove estágios comuns (Figura 21), porém com ordens e posições não absolutas, podendo ser iterativas, ocorrerem em paralelo, ou ainda serem ignoradas (Pepe et al., 2010).

É importante destacar que o modelo foi desenvolvido para aplicação em uma comunidade de investigação específica – sensores incorporados em rede. Portanto, a narrativa das etapas abaixo são direcionadas a esta área, combinando exemplos de práticas das ciências, das tecnologias e da associação entre as tecnologias e ciências

Figura 21

Modelo do ciclo de vida do CENS



Nota. Wallis et al (2008).

A combinação das práticas entre as áreas constantes do modelo são identificadas abaixo pelas respectivas letras que precedem o nome dos domínios, como segue: (a) para as ciências, (b) para as tecnologias e (c) para a associação entre as tecnologias e ciências, que representam o tipo de ação em que cada área realiza na etapa especificada:

1) **Experiment Design** (Projeto experimental): representa o início do ciclo de vida dos dados, pode constar (a) a geração de hipóteses e desenvolvimento de métodos, (b) o planejamento da amostragem e a escolha de equipamentos, (c) a negociação das questões da investigação, de pessoal e da agenda de tarefas;

2) **Calibration and Setup** (Calibração e configuração): são os ajustes de equipamentos no laboratório e em campo, isto envolve (a) regular os equipamentos, recolher as amostras, (b) preparar modelos, dados ou algoritmo a serem usados e (c) ajustar os sistemas de detecção e verificação;

3) **Capture or Generation** (Captura ou geração): os dados são produzidos ou coletados através de (a) amostragem manual, observações, processamento de amostras e (b) sensores, bem como modelos e execução de algoritmos. (c) Ajustes de sistemas e *check-in* em grupos são realizados;

As três etapas seguintes formam um subprocesso que prepara para a fase de análise.

4) **Cleaning** (Limpeza), 5) **Integration** (Integração), 6) **Derivation** (Derivação).
Limpeza: após a captura, informações sobre os dados são acrescentados para a normalização dos equipamentos. Integração: procura-se tendências, os conjuntos de dados são integrados por vários investigadores em múltiplas combinações. Derivação: para interpretar as amostras, os dados são calculados antes da análise. As práticas das três etapas podem ser vistas também na (a) análise de amostras, comparação com modelos ambientais e remoção de valores discrepantes, (b) na depuração e verificação de relatórios de erros e reteste; e, (c) na remoção dos artefatos do sensor, derivação e agregação dos dados;

7) **Analysis** (Análise): após o tratamento dos dados, usa-se ferramentas estatísticas, também de modelagem e de visualização, para testar e tirar conclusões sobre os dados, em práticas de (a) regressão linear das variáveis capturadas e teste de hipóteses, (b) pode-se ainda realizar comparações, avaliações e (c) visualizar as hipóteses testadas;

8) **Publication** (Publicação): os dados analisados culminam em publicações acadêmicas, como artigos, documentos de conferências, pôsteres e relatórios técnicos, realizado com (a)(c) a publicação das conclusões em periódicos de ciências, poster ou repositório de dados, (b)(c) procedimentos técnicos e postagem dos dados; e

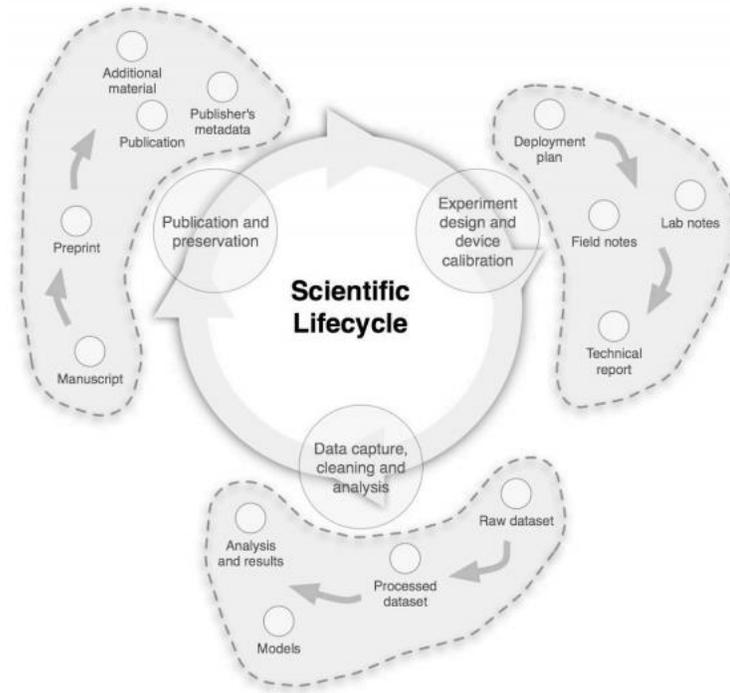
9) **Preservation** (Preservação) – são estratégias para a preservar os dados na forma de (a)(c) amostras mantidas em bancos de dados, impressos ou arquivados, mas também na (a)(c) disponibilização dos arquivos em um servidor de laboratório ou servidor local.

Pepe et al. (2010), propuseram integrar o modelo CENS com o processo de investigação. Para tal, as etapas do ciclo de vida foram condensadas em três estágios: 1) projeto experimental e calibração do dispositivo; 2) captura, limpeza e análise dos dados; e, 3) publicação e preservação (Figura 22). Portanto, é necessário visualizar as etapas do ciclo da vida dos dados como três subprocessos, ligados entre si através de um processo em cadeia relacionado ao ciclo da investigação. Ou seja, o modelo pode ser visto como uma estrutura que envolve três processos sequenciais.

O uso dos subprocessos verificados no modelo depende das características de cada investigação. Assim através das etapas e dos subprocessos, o modelo propõe uma adaptação à investigação que está a ser realizada, conforme a figura que segue.

Figura 22

Integração do modelo CENS com o processo de investigação



Nota. Pepe et al. (2010).

O estado real da investigação e da RDM foi a base para a fundamentação do modelo CENS. Neste sentido, pode-se perceber uma ênfase no ciclo de vida da investigação e na tentativa de realizar uma adaptação às características heterogêneas na gestão dos dados sob o contexto de áreas científicas diferentes. Nesse sentido, o modelo é relevante para caracterizar os ciclos de vida em áreas distintas, como é o caso da universidade. Isto é refletido pelo uso de vários subprocessos e na proposta de composição das etapas com ordens e posições a ocorrer de forma iterativa, em paralelo, ou mesmo sendo ignoradas. Estas ênfases podem fazer com que o modelo seja adaptável à investigação, e não o inverso. Também se observa que, por ter sido concebido para uma comunidade particular, etapas muito específicas são apresentadas, como é o caso da calibração e derivação. Constata-se ainda, que a ligação entre a primeira e a última etapa não é óbvia, pois não há visualização de novas ideias ou reutilização dos dados que podem reiniciar o ciclo. No entanto, seus fundamentos podem ser observados na aplicação da representação de subprocessos. Por isso permite-se inferir que o modo flexível e adaptável como proposto pelo modelo CENS, é capaz de ser acrescentado como uma característica fundamental ao modelo para a RDM aqui proposto.

4.2.9. IWGDD - INTERAGENCY WORKING GROUP ON DIGITAL DATA

Representando quase trinta agências federais nos Estados Unidos, o *Interagency Working Group on Digital Data* (IWGDD) foi estabelecido em 2006 com o objetivo de desenvolver e promover a implementação de um plano estratégico para cultivar a estrutura aberta e interoperável para as áreas de educação, ciências, tecnologia e engenharia. O IWGDD se propõe a garantir, de modo confiável, a preservação e o acesso eficaz aos dados digitais para a investigação, no intento de possibilitar o progresso da ciência e da sociedade. Neste sentido, o grupo elaborou um documento de referência (IWGDD, 2009) onde é fundamentado a RDM e enfatizado que os dados devem ser gerenciados de acordo com um plano que cubra totalmente o seu ciclo de vida que, de forma prática, possa oferecer suporte ao processo da investigação. Segundo Shepanek (2011), esta junção de finalidades do IWGDD faz parte do contexto chave para a aplicação da RDM no ambiente (Figura 23).

Figura 23

Modelo do ciclo de vida do IWGDD



Nota. IWGDD (2009).

O modelo circular do IWGDD possui uma série de práticas para a RDM e para a curadoria que permeiam o ciclo de vida dos dados. Estas práticas estão distribuídas em sete camadas com quatro ações sequenciais: *plan*, *create*, *keep* e *disposition*, mais quatro ações que ocorrem em todas as fases do ciclo de vida dos dados: *document*, *organize*, *protect* e

access e seis questões contextuais a serem consideradas: *culture, use & reuse e strategy, technical requirements, human resources & professional skills e organizations and entities, policy*. Os dados passam necessariamente pela função sequencial do ciclo de vida, no entanto podem mover-se para outras camadas à medida que são usados por diferentes entidades ou para diferentes propósitos. O IWGDD (2009), fornece as seguintes explicações sobre as etapas e as práticas compostas nas ações sequenciais:

1) **Plan** (Planejar): determina quais dados precisam ser criados ou coletados para apoiar uma agenda de investigação, identifica e avalia as fontes e os padrões para os dados, bem como os formatos de metadados, também especifica as ações e responsabilidades para realizar a RDM ao longo do ciclo de vida;

2) **Create** (Criar): produz ou adquire dados para os fins pretendidos, são criados de acordo com as finalidades definidas, os recursos de apoio para o acesso, incluindo os produtos derivados, resumos, agregações, relatórios e publicações;

3) **Keep** (Manter): organiza e armazena os dados para apoiar a investigação e para atender as necessidades especificadas no planejamento, realiza práticas de atualizações e adições às coleções existentes, de modo a garantir que os dados sobrevivam intactos pelo tempo que for necessário, assegurando sua preservação; e

4) **Disposition** (Disponibilizar): é a estratégia para a finalização do processo, com planejamento para transferir os dados para o repositório ou qualquer outra base mais adequada, ou ainda a realização do descarte dos dados.

As próximas explicações, referem-se à quatro práticas que ocorrem em todas as fases do ciclo de vida, nomeadas de funções de gestão:

1) **Document** (Documentar): define padrões para o conteúdo dos dados, sua forma, a qualidade, frequência de atualizações, entre outros. Também cria e mantém os metadados; documenta o histórico dos dados; anomalias; lacunas e registros de decisões e ações de disponibilização;

2) **Organize** (Organizar): projeta e implementa a arquitetura, a engenharia e as estruturas para os dados em conformidade com os padrões adotados;

3) **Protect** (Proteger): mantém e monitora a segurança do sistema de armazenamento e dos ativos armazenados, implementa métodos para garantir e verificar o controle da qualidade dos dados, restrições de acesso, direitos de propriedade, privacidade e confidencialidade, garantindo a autenticidade dos dados e a confiabilidade do sistema; e

4) **Access** (Acessar): garante a identificação coerente dos dados distribuídos, também fornece informações sobre os dados disponíveis, disponibilizando suporte a métodos de descoberta, análise, reaproveitamento, disseminação e apresentação dos dados.

Na camada central e nas camadas superiores do modelo IWGDD, podem ser identificadas as fases que remetem aos assuntos a serem tratados no ciclo de vida. São questões contextuais e organizacionais sobre a gestão dos dados, dispostas em etapas que envolvem a cultura, as estratégias, os requisitos técnicos, os recursos humanos, os perfis profissionais, as políticas e as instituições e entidades. Nesse sentido, as questões abaixo são exemplos sugeridos pelo autor para discussões que podem ser realizadas nestas camadas, resultando em ações de resolução de problemas e aplicação para a RDM.

- a) A cultura como uma barreira ou ativo potencial para a partilha e uso dos dados;
- b) O desenvolvimento dos recursos humanos que permitam a aquisição de habilidades para a compreensão e a realização da RDM;
- c) A necessidade de estruturas organizacionais suficientemente ágeis para abordarem os desafios da ciência, mas também formais para serem eficazes na RDM;
- d) A implementação de políticas estruturadas e eficazes para o apoio ao empreendimento científico, de forma ampla; e
- e) A necessidade de centralização da atenção às tomadas de decisão relacionadas aos requisitos técnicos, como a retenção, preservação e transferência dos dados, ou o entendimento de que os planos de gestão de dados podem variar para cada ambiente da RDM.

Como exemplo de uso dessas camadas, a IWGDD propõe que nas etapas três (proteger) e quatro (acessar) do ciclo sequencial, sejam realizadas ações para adquirir, atualizar e implementar as tecnologias, de modo a superar a obsolescência e melhorar o desempenho ou expandir a capacidade de armazenamento e processamento, implementando o apoio para as necessidades de evolução, análise e acesso aos dados.

No modelo apresentado pelo IWGDD pode-se perceber que a proposta excede o ciclo de vida dos dados e da investigação, incorporando elementos contextuais, como, por exemplo, a própria instituição ou a preocupação com a parte técnica, ou ainda com a política normativas dos dados, mas também, com as potencialidades humanas e o preparo profissional para gestão dos dados. Assim, o modelo abarca práticas para a RDM e para a curadoria. No entanto, em termos visuais, é de difícil entendimento, especialmente por estar disposto em um modelo circular, o que remete a necessidade de produzir um documento que especifica a intencionalidade do diagrama de forma clara. Este modelo, assim como o do DCC contribuem para mostrar que o ciclo de vida dos dados e da curadoria, fazem parte de um contexto amplo do ambiente da RDM, que envolve diversos aspectos a serem considerados e detalhados, o que se assemelha em intencionalidade ao modelo para a RDM que é proposto neste trabalho.

4.3. SÍNTESE DOS MODELOS DE CICLOS DE VIDA

É possível sintetizar os modelos de ciclo de vida, seus elementos e processos apresentados ao longo deste capítulo, através da seleção e organização do conjunto de etapas, características e peculiaridades estudadas, com o objetivo de reconhecer quais são as práticas essenciais para a RDM que possibilitam uma partilha adequada dos dados. Tal sintetização é realizada nesta seção da seguinte forma: primeiro, identificam-se as práticas recorrentes nas etapas sequenciais dos modelos (Tabela 18); depois, agrupam-se e caracterizam-se estas etapas identificadas, donde resulta a sintetização das práticas sequenciais recorrentes. Posteriormente, os elementos relevantes que envolvem as práticas sequenciais, ou seja, as práticas contextuais, são identificadas a partir do contexto de cada modelo e das etapas importantes para a conjuntura estudada (Tabela 19), produzindo o agrupamento e a caracterização das práticas nomeadas de contextuais.

A sintetização e a proposição do agrupamento das práticas sequenciais e das práticas contextuais de modo conjunto, produzem as práticas essenciais para a RDM (Figura 25). Portanto, de modo a melhor esclarecer a finalidade destes agrupamentos de etapas e conseqüentemente evitar a confusão entre eles, segue a definição e a diferenciação do significado das práticas nomeadas de contextuais e das práticas nomeadas como sequenciais.

As práticas contextuais são etapas que trazem consigo aspectos relevantes da RDM e da curadoria, cujas ações abrangem e ultrapassam o contexto do processo de investigação. As práticas deste tipo ocorrem em mais de uma fase do modelo, ou seja, são elementos cujos objetivos excedem a própria etapa e podem ser realizadas também em parte ou em todas as outras etapas do ciclo de vida. Por exemplo, a etapa nomeada de “preservação”, em alguns modelos é representada no final do ciclo. Porém, o seu mote não se restringe a uma etapa, na verdade é um objetivo que deve ser considerado desde o momento do planejamento, passando pela coleta e indo até o final do processo. Assim, é essencial para a RDM que esta fase possa ser distribuída em várias etapas do modelo, o mesmo raciocínio pode ser aplicado às outras fases como, por exemplo, a privacidade e os direitos autorais (Sant’Ana, 2013).

As práticas sequenciais, diferente das etapas contextuais, são exercidas predominantemente nas fases que fazem parte do fluxo sequencial constante nos modelos. Normalmente relacionadas ao processo de investigação, são caracterizadas por ações bem definidas, realizadas quase que exclusivamente no momento determinado pela etapa, como por exemplo, a coleta dos dados. Portanto, as suas ações são basilares para as práticas contextuais e repercutem de modo fundamental para a curadoria e para a RDM, bem como encontra o seu apoio no processo de investigação.

Primeiramente segue abaixo a sintetização proposta para as práticas sequenciais, através da seleção das práticas recorrentes; formação do agrupamento e caracterização das etapas. Neste sentido, a partir da análise dos nove modelos aqui estudados, a tabela elaborada a seguir apresenta as etapas recorrentes e a proposta do agrupamento das etapas predominantes. O agrupamento é realizado pelas cores comuns, ou seja, as fases que se mostram conceitualmente semelhantes estão identificadas pela mesma cor. A intencionalidade de cada etapa é expressa através dos verbos que a denomina, sendo apresentada em ordem numérica, de acordo com a sequência constante nos modelos. A legenda da tabela nomeia a proposta de agrupamento identificado pela sua ideia principal, segue após a tabela a caracterização das etapas agrupadas, para posterior organização visual (Figura 24), resultando nas práticas sequenciais a serem utilizadas para a RDM no propósito de uma partilha adequada dos dados.

Tabela 18*Etapas sequenciais recorrentes nos modelos de ciclo de vida*

MODELO	ETAPAS							
	1	2	3	4	5	6	7	8
DCC	Conceituar	Criar e receber	Avaliar e selecionar	Arquivar	Preservar	Armazenar	Aceder, usar e reusar	Transformar
DDI	Conceituar	Coletar	Processar	Arquivar	Distribuir	Descobrir	Analisar	Reaproveitar
e-Science	Conceituar e projetar	Coletar	Processar	Acessar e disseminar	Analisar	Transferir	Finalizar	Reaproveitar
Jisc	Planejar	Coletar e capturar	Colaborar e analisar	Armazenar	Partilhar	Descobrir	Reutilizar	
DataONE	Planejar	Coletar	Assegurar	Descrever	Preservar	Descobrir	Integrar	Analisar
CENS	Projetar	Calibrar e configurar	Capturar ou gerar	Limpar, integrar e derivar	Analisar	Publicar	Preservar	
UVa	Planejar	Iniciar	Coletar	Analisar	Partilhar	Finalizar		
UKDA	Criar	Processar	Analisar	Preservar	Acessar	Reutilizar		
IWGDD	Planejar	Criar	Manter	Dispor				

● *Planejar*
● *Recolher*
● *Analisar*
● *Partilhar*
● *Reutilizar*
● *Descrever*

● **Etapa 1 - Planejar:** todos os modelos estudados apresentam na primeira etapa algum tipo de planeamento para o ciclo de vida, porém tal concepção mostra-se abrangente, vai desde a ideia inicial para a questão da investigação (DDI), que pode ser via descoberta dos dados (UVa), até o plano do projeto da investigação (UKDA), passando pela adoção de um plano de gestão dos dados (DataONE), bem como o planeamento das opções de armazenamento (DCC), dos princípios FAIR, das políticas, serviços e requisitos relevantes para o contexto organizacional (Jisc), incluindo a negociação de recursos humanos (IWGDD) e outros recursos para a investigação (CENS). Há também a previsão de um arranque para o plano (UVa), onde as atividades necessárias são realizadas antes do início do projeto. Esta etapa destaca fundamentalmente a importância do plano de gestão de dados, que contém em seu escopo a maior parte das preocupações observadas nos modelos.

● **Etapa 2 – Recolher:** é a coleta, captura ou geração dos dados que pode igualmente ser visualizada em todos os modelos, em oito deles na segunda fase. São os dados primários e secundários (DDI) necessários para conduzir a investigação (UVa), gerados pelos

investigadores através de equipamentos (CENS), ou coletado por outras fontes, como os repositórios (DCC). Nesta etapa podem existir ações para a elaboração da documentação e do armazenamento dos dados coletados (DataONE), bem como o cuidado com os metadados (Jisc). A UKDA usa a sua primeira etapa para abarcar tanto o planeamento quanto a coleta, já o CENS constitui na sua segunda etapa a preparação ferramental para a coleta. A fase da recolha é comum aos modelos estudados, sendo caracterizada por outras ações que ocorrem em paralelo, não sequenciais, mas essenciais para a RDM e para a curadoria, como por exemplo, a preservação e a reutilização dos dados.

● **Etapa 3 – Analisar:** esta etapa pode ser percebida nos modelos de duas formas, como análise dos dados digitais para a curadoria e preservação (DCC), ou como análise necessária para desenvolvimento dos objetivos da investigação (UVa), abrangendo também outros tipos de dados que não sejam digitais (DDI). Assim, o DataONE usa o termo “Assure” para incluir a análise de forma a garantir a qualidade dos dados, o Jisc concentra a análise dos dados para a realização da partilha, a UKDA enfatiza a análise para o armazenamento, enquanto o CENS usa a etapa da análise direcionada à investigação. O objeto principal desta etapa é a adoção da “análise” para a RDM e para as ações referentes a curadoria dos dados recolhidos, porém esta fase é indissociável da análise realizada e constante do processo de investigação.

● **Etapa 4 – Partilhar:** o ato da partilha pode ser destacado pela partilha realizada entre os investigadores durante as fases de recolha e análise, ou formalmente quando da distribuição de forma pública e geral após finalizada a investigação (Jisc). Em ambos os casos, a partilha dos dados deve ser realizada necessariamente por ferramentas distribuídas e seguras, pois minimiza as perdas potenciais em curto prazo e em longo prazo (DataONE). A partilha após a análise é enfatizada nos modelos estudados, de modo que os dados possam ser disponibilizados em repositórios, centros de dados e destinados à preservação via armazenamento seguro (DCC), resultando na proteção dos dados e metadados e na agregação de valor ao longo do tempo (DDI), incluindo nesta etapa as considerações legais e técnicas para proporcionar segurança (Jisc). Deve-se considerar, ainda, os cuidados pertinentes com as atualizações e adições às coleções existentes (IWGDD), e a partilha formal por meio das apresentações, publicações e outros produtos da comunicação científica (UVa).

A ideia geral da partilha é que ela não se resume ao ato, mas constitui uma série de ações que remetem aos objetivos comuns da RDM e da curadoria como, por exemplo, arquivar, preservar, armazenar, publicar, distribuir ou descartar, junto com as interações técnicas e legais para garantir a proteção, confiabilidade e segurança dos dados. Aqui também é destacada a necessidade dos recursos humanos e de outros recursos no apoio ao investigador.

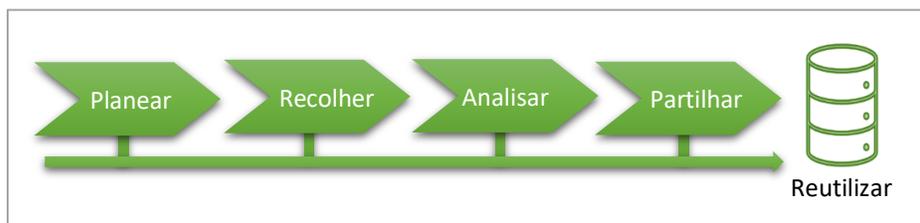
○ **Etapa 5 – Reutilizar:** as práticas relacionadas para a reutilização dos dados garantem que eles estejam acessíveis aos usuários após o término da investigação, ou seja, para as comunidades envolvidas com a curadoria e aos interessados na reutilização dos dados (DCC), de modo que possam ser utilizados em uma estrutura conceitual diferente da originária (DDI), ou gerar novas ideias para projetos de investigação e assim possibilitar um novo início do ciclo de vida (UVa). Esta etapa pode conter finalidades diferentes para atores diversos, a reutilização dos dados permite ser percebida sob duas perspectivas: na gerência do processo e para o consumo dos dados (Jisc). Na gerência do processo, existem ações de promoção para a descoberta dos dados, de modo que o ciclo não termine na partilha. Já o consumo dos dados pode ser visto para o reaproveitamento na própria investigação ou por terceiros, com ações de procura de fontes externas para a recolha e reutilização dos dados em novas ideias para a investigação. Porém, a reutilização para a RDM envolve aspectos mais abrangentes que podem repercutir na ciência e na sociedade. Nesse sentido, é necessário organizar todo o processo da RDM para permitir que os dados sejam assim reutilizados.

○ **Descrever:** como uma prática fundamental para a RDM, a etapa da descrição atribui os tipos de metadados pertinentes e necessários para a compreensão contextual e para a apresentação do dado digital, utilizando padrões apropriados para garantir uma descrição adequada e o controle em longo prazo (DCC). Embora seja destacada como uma etapa sequencial pela DataONE, a descrição pode ser considerada um elemento contextual do ciclo de vida, pois é pertinente que as suas práticas estejam disponíveis para serem realizadas em outras fases do ciclo de vida, ou seja, de modo não sequencial. Uma minuciosa e completa caracterização sobre os dados de investigação é realizada aqui por meio da adoção de padrões e ferramentas de metadados. Portanto, essenciais para a partilha e na reutilização dos dados, esta etapa será considerada nesta investigação como uma prática contextual, pelo que não estará no conjunto de práticas sequenciais apresentados a seguir.

O resultado da seleção e do agrupamento das práticas sequenciais, a partir dos modelos estudados, podem ser visualizados na figura abaixo, constando cinco etapas representadas, quais sejam: 1) planear; 2) recolher; 3) analisar; 4) partilhar; e 5) reutilizar. Como já mencionado, uma etapa citada mas não representada é a etapa “descrever”, que fará parte das práticas contextuais, pois mostra-se importante de modo contextual nos ciclos de vida estudados. Portanto, com base na seleção, agrupamento e caracterização apresentadas anteriormente, as práticas sequenciais aqui resultantes e, a serem utilizadas para a RDM, no propósito de realizar uma partilha adequada dos dados, fazem-se retratadas e sintetizadas na figura abaixo.

Figura 24

Práticas sequenciais a serem utilizadas para a RDM em uma partilha dos dados



Por sua vez, as práticas contextuais – segundo agrupamento de etapas que formam as práticas essenciais – são organizadas tomando por base os aspectos relevantes das etapas não sequenciais constantes nos modelos, normalmente, caracterizadas pelas ações para RDM e para a curadoria. Conforme definido no início desta seção, as ações que constituem as etapas contextuais, abrangem e ultrapassam o ambiente do processo de investigação, podendo ocorrer em apoio e associação com mais de uma fase das etapas sequenciais, agora organizada, apresentadas e representadas anteriormente na figura 24.

No sentido exposto acima, a tabela seguinte reúne as sentenças que retratam os aspectos do contexto dos modelos estudados, juntamente com as etapas do contexto também identificadas como não sequenciais. Ou seja, os elementos contextuais são percebidos a partir do ambiente envolvente dos ciclos de vida, das práticas sequenciais e das etapas para a curadoria e para a RDM na conjuntura estudada. Cabe ressaltar, que os aspectos do contexto presentes na tabela 19, contêm as características gerais do modelo estudado, inclusive da estrutura, para identificar o que seja relevante e ajudar na concepção do modelo final aqui proposto. Após a tabela, são caracterizadas as etapas contextuais já agrupadas.

Tabela 19*Elementos contextuais nos modelos de ciclo de vida*

Modelos	Elementos contextuais: aspectos e etapas dos modelos
DCC	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → indicativo em vez de exaustivo e adaptável a diferentes domínios; → usa-se em conjunto com as normas para planejar as atividades de curadoria; → a operacionalização depende da necessidade do contexto da organização ou domínio; → permite a entrada para uso em qualquer estágio do ciclo de vida; → ajuda a definir os fluxos de trabalho, funções e responsabilidades a ela associadas; → oferece suporte em uma abordagem holística; → o dado digital e os bancos de dados encontram-se como elementos centrais; → constitui-se por ações realizadas durante todo o ciclo de vida, sequenciais e ações realizadas conforme as circunstâncias exigem; → acompanha as atividades das comunidades envolvidas com a curadoria; → realça a participação no desenvolvimento das ferramentas para partilha e curadoria; → destaca a preservação e o cuidado com os metadados; e → reavalia periodicamente os objetos digitais. <p><i>Etapas:</i> curar e preservar, participar e acompanhar, descrever e representar.</p>
DDI	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → circular, embora em parte mostre-se linear com um caminho alternativo e um loop; e → simples, com aspectos comuns de fácil visualização e associação à investigação; <p><i>Etapas:</i> distribuição e descoberta dos dados.</p>
e-Science	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → proporciona um melhor entendimento das relações entre as etapas da investigação; → aumenta a consciência das perdas potenciais de informação entre as fases; → define como pontos de transição os locais mais vulneráveis à perda de informações, ocorrendo à medida que se passa de uma etapa para a seguinte; → pontos de transição incluem também as atividades da RDM e das partes interessadas; → demonstra a necessidade da transferência de conhecimento; e → indica que o arquivo dos dados e os resultados da investigação não é o fim do processo. <p><i>Etapas:</i> transferência de conhecimento</p>
UVa	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → demonstra com clareza a separação e a identificação das etapas que envolve o processo de investigação, a curadoria e os outros componentes que estão no ambiente; → disponibiliza recursos que acompanham os modelos para facilitar a identificação e uso; e → identifica a possibilidade de geração de novas ideias com o reinício do fluxo.
DataONE	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → a estrutura orienta a utilização de recursos, serviços e materiais da área; → as etapas podem ser expandidas, possibilitando novos fluxos; → dispõe de um guia de melhores práticas para gerenciamento dos dados; → nem todas as etapas do ciclo de vida segue um caminho sequencial, mas depende do tipo de uso a ser realizado; → pode combinar vários e diferentes fluxos de trabalhos em uma expressão circular; e → a maior ênfase são nos dados e em como gerenciá-los <p><i>Etapas:</i> assegurar, descrever, preservar</p>

Jisc	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → ênfase na RDM e na curadoria digital; → envolvido por práticas principais ou preocupações para a RDM; → permite inferir a necessidade de atores diferentes para a execução das ações propostas; → apresenta uma disposição organizacional fundamentada para que as práticas sejam implementadas; e → exibe uma lista de checagem do modelo que permite conhecer as práticas e destacar as ferramentas para execução de cada fase do ciclo.
	<p><i>Etapas:</i> colaborar, gerenciar ativos de dados, armazenar e preservar, catalogar e registrar</p>
UKDA	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → visa ajudar a associar as práticas para a RDM ao processo da investigação; → direciona as atividades de suporte, orientação e serviços para a RDM; → centrado em dados e no processo de investigação; → os recursos não têm forte ênfase no modelo do ciclo de vida; → dificuldade de vincular o investigador com o processo; → baixo nível de abstração, praticamente fechado, unidirecional e serial; e → pode haver uma interrupção entre etapas em termos de reinício do ciclo de vida.
CENS	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → elaborado a partir do estado real da investigação; → desenvolvido para aplicação em uma comunidade de investigação específica; → combina práticas das ciências, tecnologias e da associação entre tecnologias e ciências; → estrutura circular que envolve três subprocessos sequenciais; → ciclo de vida com três subprocessos ligados entre si em um processo em cadeia relacionado ao ciclo da investigação; → proporciona adaptação à investigação que está a ser realizada; → ordens e posições das etapas não absolutas e iterativas, podendo ocorrer em paralelo, ou serem ignoradas; → ênfase no ciclo de vida da investigação e adaptação à heterogeneidade da RDM; e → não há visualização de novas ideias ou reutilização dos dados para o reinício do ciclo.
	<p><i>Etapas:</i> preservar</p>
IWGDD	<p><i>Aspectos:</i></p> <ul style="list-style-type: none"> → possui ações sequenciais, ações que ocorrem em todas as fases do ciclo de vida dos dados e ações contextuais; → Os dados podem mover-se para outras camadas à medida que são usados por diferentes entidades ou para diferentes propósitos; e → mostra que a o ciclo de vida dos dados e da investigação fazem parte de um contexto amplo do ambiente da RDM.
	<p><i>Etapas:</i> documentar, organizar, proteger, acessar, cultura, estratégias, requisitos técnicos, recursos humanos e perfis profissionais, políticas, instituições e entidades.</p>

As etapas apresentadas na tabela anterior, caracterizam-se por envolverem o ambiente dos ciclos de vida e as práticas sequenciais, bem como são orientadas para as ações de curadoria e para a RDM. Assim, tendo por consideração a relevância com que estas fases se apresentam nos modelos, segue uma proposta de agrupamento das etapas constantes na tabela, resultando nas práticas nomeadas como contextuais.

• **Proteger:** refere-se à proteção dos dados durante o projeto, incluindo as considerações legais e técnicas (Jisc), mas também se preocupa com os aspectos de segurança dos sistemas, restrições de acesso, direitos de propriedade, privacidade e confidencialidade, garantindo a autenticidade e a segurança dos dados, bem como a confiabilidade dos sistemas (IWGDD). Pelos muitos aspectos que podem ser aqui detalhados, esta etapa abrange diversas características a serem aplicadas junto com as práticas sequenciais do modelo.

• **Estruturar:** realizam-se aqui, ações de modo a prover, uma atenção fundamental para a implementação e uso da arquitetura, da engenharia e para a organização dos dados com métodos e controle de qualidade em uma ciberinfraestrutura distribuída que visa atender as necessidades apresentadas (DataOne). Utilizam-se para tal, os recursos humanos e perfis profissionais necessários (IWGDD), permitindo a participação nas atividades em que os artefatos adequados à curadoria possam ser desenvolvidos e partilhados (DCC), de modo a possibilitar que os colaboradores reutilizem os dados sem dificuldades, com plataformas de colaboração e comunicação pertinentes, gerindo os dados com os cuidados necessários (Jisc). Todas as fases e modelos dependem, de alguma forma, de uma estrutura ou infraestruturas físicas e tecnológicas para serem realizadas em seus contextos.

• **Monitorar:** é o acompanhamento intencional das ações para a gestão, planejadas para promover a curadoria por todo o ciclo de vida (DCC), assistindo as estratégias políticas e os relacionamentos com instituições e outras entidades (IWGDD), bem como preocupando-se com as perdas de informações e as transferências de conhecimento durante o ciclo de vida, de modo a evitar eventuais prejuízos aos atuais e futuros investigadores no processo para a utilização e reutilização dos dados (*e-Science*). Embora esta etapa não esteja de modo tão perceptível quanto as outras fases no ciclo de vida, torna-se útil, pois o contexto das etapas requer um tipo de atenção contínua para as suas ações, de modo que interlocuções sejam realizadas e as práticas de curadoria sejam acompanhadas;

• **Preservar:** está no contexto de todos os modelos estudados e especificado por quatro práticas sequenciais. A preservação é importante por considerar, entre outros elementos, o local de armazenamento (CENS) e por atentar rotineiramente para a salvaguarda dos dados (Jisc), com a elaboração de planos de preservação para as ações

realizadas ao longo do ciclo de vida (DCC), mas também para orientar o envio dos dados (DCC), metadados (DDI) e outros objetos digitais para um local apropriado que possa preservá-los em curto e em longo prazo, de modo que permitam às partes interessadas acessar, interpretar e usar os dados no presente e futuro (DataONE).

- **Perdas de dados:** é o alerta para diminuir as perdas potenciais de informação, no caso deste estudo, os dados. São lacunas identificadas entre as etapas mais suscetíveis às perdas, definindo-as como pontos de transição ao longo do ciclo de vida. Ocorre à medida da passagem de uma etapa para outra, caracterizando-se como elementos de atenção para o modelo e necessários para a transferência de conhecimento. Esta etapa está fundamentada no modelo da *e-Science*. Embora não esteja de forma recorrente nos contextos dos modelos e das etapas apresentadas, esta fase é incluída como uma prática essencial pela peculiaridade e importância, quando objetivada a partilha dos dados e a sua consequente reutilização, pois considera-se que a minimização das perdas informacionais e consequentemente dos dados, é essencial para que este e outros objetivos sejam cumpridos.

Após a seleção e caracterização das práticas sequenciais e contextuais, é necessário unir estas duas seleções para que as práticas essenciais estejam identificadas e visualizadas em uma ordem lógica e relacional. Assim, para esta a sintetização e visualização, a figura 25 reúne as etapas selecionadas e identificadas como essenciais para a RDM no objetivo da partilha adequada dos dados, após a figura são explicados os relacionamentos para as práticas essenciais.

Figura 25

Práticas essenciais a serem observadas para a RDM



A figura apresentada indica que o encadeamento das cinco **práticas sequenciais** – planejar, recolher, analisar, partilhar e reutilizar – estão envolvidas por vários itens necessários para que uma partilha dos dados aconteça. Por sua vez, estes itens envolventes nomeados de **práticas contextuais**, estão agrupados em cinco etapas – descrever, proteger, estruturar, monitorar e preservar. As práticas contextuais são sustentadoras e estão ligadas para acontecer a qualquer momento e no decorrer das práticas sequenciais. A conjugação destes elementos tendo por base a organização, os fundamentos e as características já apresentadas, formam as **práticas essenciais**. Portanto, este conjunto de práticas assim relacionadas, podem direcionar para que a RDM e a partilha dos dados, seja realizada de forma adequada no ambiente da universidade.

Por fim, uma outra sintetização possível de ser concebida nesta seção, refere-se às observações que são frutos da análise dos modelos ao longo do capítulo. Ou seja, como resultado da análise de cada um dos nove modelos de ciclos de vida, as conclusões provenientes das suas estruturas e características, servem para ratificar os propósitos aqui concebidos nas práticas essenciais e contribuem, juntamente com a tabela anterior, para orientar a estrutura do modelo final desta tese. Portanto, as conclusões, peculiaridades e as características observadas nos modelos estudados, seguem sintetizados na tabela abaixo.

Tabela 20

Contribuições para o desenvolvimento do modelo para a RDM

Modelos	Contribuições para o desenvolvimento do modelo para a RDM
DCC	<ul style="list-style-type: none"> • A ideia de ações ocasionais, viabilizadas conforme as necessidades sejam apresentadas; • Práticas sequenciais e o processo de investigação adaptadas à outras partes do modelo; • Uso de algoritmos e fluxos de trabalho de investigação integrados por softwares; • Adaptabilidade e flexibilidade no ambiente de uso, para que o modelo, ou partes dele, possam ser implementados para a RDM; • Fundamentos da ciência aberta e do acesso aberto usados em uma RDM organizacional; e • A ideia de que as ações não podem ser realizadas sem a colaboração e responsabilidades dos atores que influenciam e atuam nas etapas do ciclo de vida de dados.
DDI	<ul style="list-style-type: none"> • Aspectos comuns à investigação, tornam mais fácil a visualização e a associação pelos usuários; e • Simplicidade da ação sequencial e dos fluxos alternativos.
<i>e-Science</i>	<ul style="list-style-type: none"> • Necessidade da transferência de conhecimento; e • Idealização de pontos de transição como marco de atenção para as vulnerabilidades que ocasionam as perdas de dados.

UVa	<ul style="list-style-type: none">• Clareza com a separação e a identificação das etapas que envolvem o processo de investigação, curadoria e outros componentes do modelo; e• Disponibilização dos recursos que devem acompanhar os modelos para facilitar sua identificação e uso.
DataONE	<ul style="list-style-type: none">• As etapas não precisam ser necessariamente sequenciais, podendo inclusive serem definidas e expandidas em fluxos auxiliares.
Jisc	<ul style="list-style-type: none">• Alerta para as atribuições de funções e responsabilidades requerida para a partilha;• Aspectos dos princípios e práticas da ciência aberta permeiam todas as fases do modelo; e• Designa outros atores para realização das práticas que excedem o trabalho do investigador no processo de investigação.
UKDA	<ul style="list-style-type: none">• Ênfase nos dados em associação direta com o processo da investigação; e• Cuidados com a falta de etapas que possam direcionar e explorar a RDM e a curadoria.
CENS	<ul style="list-style-type: none">• Aplicação da ideia dos subprocessos, permitindo um modo flexível e adaptável; e• Realização de uma adaptação às características heterogêneas do gerenciamento dos dados, proposto pelo contexto, em áreas científicas diferentes.
IWGDD	<ul style="list-style-type: none">• Mostra que o ciclo de vida dos dados e da curadoria fazem parte de um contexto amplo do ambiente da RDM; e• Necessidade de produzir um documento que especifica de forma clara a intencionalidade do modelo.

Os parâmetros das duas últimas tabelas, bem como as práticas essenciais aqui identificadas e descritas, são utilizados, norteiam e constituem o modelo para a RDM proposto, o que pode ser constatado no capítulo 7, quando os pontos norteadores para a constituição do modelo gráfico é apresentado e posteriormente seus elementos e processos são relacionados e explicados de modo textual.

4.4. CONSIDERAÇÕES SOBRE AS PRÁTICAS ESSENCIAIS PARA A GESTÃO DOS DADOS

Este capítulo apresentou nove modelos de ciclos de vida levantados na literatura científica que representam os fluxos da RDM, da curadoria e dos dados de investigação. A abordagem permitiu identificar e descrever as práticas essenciais para a RDM (Figura 25) a serem utilizadas para sustentar uma partilha adequada dos dados. Com tal resultado, a questão complementar que pretendia saber quais são as práticas essenciais para a RDM foi respondida. O caminho percorrido para a obtenção do resultado deste capítulo, primeiramente estabeleceu de modo teórico uma visão geral das diferenciações, convergências e limites entre as práticas de curadoria digital e da RDM no processo de investigação, considerando o ciclo de vida dos dados, que além de permitir obter uma visão geral sobre o tema, proporcionou um melhor entendimento para a análise dos elementos envolvidos nos modelos de ciclos de vida estudados.

Portanto, com os parâmetros firmados, a exploração dos modelos resultou na identificação das práticas sequenciais e das práticas contextuais, que, quando sintetizadas e descritas resultaram nas práticas essenciais com seus elementos e processos para a RDM, cumprindo-se então o objetivo 3 da tese. Ainda, através do estudo dos modelos de ciclos de vida foram recolhidas contribuições para concepção do modelo final do trabalho.

A compreensão das práticas essenciais que envolvem a RDM e a curadoria no processo de investigação, dentro de uma perspectiva de ciclo de vida, quando observados juntamente com os atores no ambiente da universidade, permite inferir que tal ambiente percebido nesta associação requer uma definição de “quem”, “como” e “em que momento” são realizadas as práticas essenciais aqui encontradas para a RDM. A diversidade dos termos e a multiplicidade de ações nas etapas podem naturalmente causar problemas para identificação dos atores participantes e na associação de responsabilidades para as práticas de gestão e da partilha dos dados. Um exemplo disso está nos modelos de ciclos de vida estudados, onde vários deles contemplam recursos detalhados para explicar e auxiliar na realização do modelo, mas não indicam “como” e “por quem” são realizados. Obviamente que a finalidade de muitos modelos não passa por esse objetivo, porém quando pensado no contexto mais abrangente como o da universidade, questões relativas à RDM em termos de organização dos atores de forma lógica e da distribuição de responsabilidades para as ações de curadoria, da RDM e da própria investigação, são imprescindíveis para que a utilização das práticas essenciais sejam bem-sucedidas.

Pode-se também entender que a falta de percepção do ambiente e de definições sobre tal organização lógica dos atores e distribuição de responsabilidades, é capaz de tornar a utilização das práticas essenciais para a RDM um tanto quanto confusas e imbricadas, quando observadas indistintamente sob os itens e peculiaridades que estas práticas possuem, sem uma reflexão que seja mais ampla sobre o contexto e seus diversos atores. A figura que segue, representa tanto esta possibilidade de confusão e justaposição de trabalhos, quanto permite visualizar o que foi estudado neste capítulo para a RDM em uma partilha adequada dos dados, tendo por eixo as práticas essenciais resultantes deste capítulo.

Figura 26

Estrutura da RDM, tendo por eixo as práticas essenciais



A figura 26 reforça o imbricamento entre as práticas essenciais com o ambiente para a RDM aqui proposto. Tal sobreposição perpassa de modo transversal o processo de investigação, sob as etapas do ciclo de vida dos dados e os cuidados com a curadoria digital, sustentados pelos princípios da ciência aberta. Assim, embora o objetivo de identificar as práticas essenciais tenha sido alcançado, outras questões podem permanecer abertas e confusas se não bem planeadas e adaptadas às características diversas e heterogêneas inerentes ao ambiente da universidade. Portanto, a construção e a organização realizada neste capítulo, através das sintetizações que resultaram nas práticas essenciais para a RDM de modo a sustentar uma partilha adequada dos dados, podem não ser suficientes em termos de obter uma melhor aplicação, se não for dado o relevo ao ambiente geral e ao contexto específico com os seus atores diversos. Isto porque a RDM e as suas práticas essenciais dependem e devem ser suportadas por uma organização maior que envolva e articule de forma clara os atores, mas também de um contexto próprio para a realização das práticas aqui encontradas.

Tendo por atenção o acima exposto, o próximo capítulo é desenvolvido sob a necessidade de articular os atores, no ambiente da universidade, para a RDM, considerando neste cenário que as práticas essenciais podem ser utilizadas de forma a potencializar a aplicação de tal gestão, ao mesmo tempo que deve ser observado o contexto abrangente e institucional da universidade, sem com isso desconsiderar as especificidades do ambiente da investigação para a RDM que resulte em uma partilha adequada dos dados.

5. TEORIA DA ATIVIDADE - A ARTICULAÇÃO DOS ATORES

A teoria da atividade tem aqui sua competência como uma ferramenta apropriada para contribuir com uma solução para a articulação dos atores para a RDM (*Research Data Management*), mas também é capaz de fornecer uma estrutura metodológica para minimizar os problemas da complexidade e da heterogeneidade no ambiente da universidade. Isto pode ser viabilizado através de duas ações: 1) a identificação, arranjo e articulação sistêmica dos atores; e, 2) a adaptação da atividade da RDM ao contexto da universidade. Tais perspectivas são possíveis de serem realizadas porque a teoria da atividade pode atuar como uma lente teórica, fundamentada no caráter sociotécnico, através do modelo do sistema de atividade e da análise das relações que o constitui, direcionada à aprendizagem para transformação da própria atividade e capaz de resultar em quadros interpretativos e propositivos. Em síntese, este capítulo produzirá um modelo para conceber um sistema de atividade para a RDM e parâmetros para a sua adaptação ao ambiente da investigação.

O resultado proposto justifica-se pela dificuldade da articulação entre os atores envolvidos na RDM e a tendência para a realização de práticas individuais que contribuem com a fragmentação e a invisibilidade dos dados no contexto da universidade. São problemas emergentes para a RDM, que em parte se deve ao ambiente complexo e heterogêneo da universidade, constituído por um conjunto diversificado de atores humanos e não humanos que precisam atuar em associação. Estes aspectos asseveram a necessidade de conceber uma estrutura bem delineada para associar os atores diversos, com interesses distintos em um esforço articulado para gerir e partilhar os dados de investigação resultantes de áreas disciplinares dissemelhantes.

Neste cenário intrincado e díspar, é necessário que as universidades instrumentalizem a RDM no ambiente da investigação, para que tal ambiente possa ser compreendido e organizado e os dados de investigação, através de práticas essenciais, se tornem adequadamente partilháveis e acessíveis, de modo a permitir serem reutilizados, potencializando assim o seu valor para a ciência e para a sociedade. Logo, diante do referido cenário, as perguntas que decorrem da questão principal desta tese são as seguintes: como deve ser a articulação dos atores? E como a RDM pode adaptar-se ao ambiente complexo e

heterogêneo da universidade? Estas questões complementares estão fundamentadas no pressuposto de que a articulação de modo estruturado é essencial para a comunicação, colaboração e coordenação, e que a sua vulnerabilidade ou ausência resultam em uma RDM limitada e de práticas isoladas, e ainda, sob tal complexidade, a articulação precisa ter uma estrutura para vincular os atores de acordo com o ambiente.

Este construto e contexto têm ainda sua gênese no segundo capítulo, onde na resposta da questão complementar foi lá apontado que, um caminho possível para a identificação dos atores da RDM na universidade, é a concepção de agrupamentos de atores. Contudo, por se tratar de um ambiente complexo e heterogêneo, tais atores dependem de um contexto específico a ser explorado, considerando que os elementos humanos e não humanos podem variar muito significativamente entre áreas disciplinares diferentes.

Para alcançar o resultado do presente capítulo e produzir um modelo que permita conceber um sistema de atividade para a RDM, bem como parâmetros para a sua adaptação ao ambiente, o estudo da teoria da atividade será direcionado no interesse deste trabalho em quatro tópicos principais, quais sejam: 1) os fundamentos essenciais da teoria da atividade; 2) o modelo do sistema de atividade; 3) o uso da aprendizagem expansiva; e, 4) uma proposta de modelo para sistematização dos atores e adaptação contextual. Assim, o tópico dois permite identificar, dispor e articular os atores no modelo sistêmico. O tópico três apresenta a principal ferramenta conceitual para adaptação à heterogeneidade do ambiente da universidade. O quarto tópico é o produto dos tópicos anteriores, e todos os tópicos são fundamentados pelo tópico um.

5.1. FUNDAMENTOS ESSENCIAIS DA TEORIA DA ATIVIDADE

A teoria da atividade foi concebida a partir da escola russa, tendo por base conceitos Marxistas (Davydov & Radzikhovskii, 1985) e articulada no trabalho do psicólogo russo Lev Vygotsky (1978, 1986), principalmente na perspectiva da análise do desenvolvimento do indivíduo. Aleksandr Luria e Aleksei Leont'ev, ao incorporarem os elementos da sociedade, cultura e dimensões históricas aos estudos de Vygotsky (Eilam, 2003; Stetsenko, 2003), deram origem à escola nomeada posteriormente de teoria da atividade histórico-cultural (CHAT, do inglês *Cultural-Historical Activity Theory*) (Cole & Engeström, 1993; Daniels, 2001; Engeström, 2014).

Da tradição russa até suas formas atuais, a teoria da atividade foi expandida para o ocidente através da academia anglo-saxônica (Engeström, 2014; Lave, 1993; Scribner, 1985), ocorrendo um crescimento exponencial em sua utilidade. Nesse sentido, deu origem a múltiplas variações em diversas disciplinas e sob diferentes abordagens (Roth, 2004). Cole (1996) e Roth e Lee (2007) atribuem essa diversidade, em parte, à capacidade da CHAT de focar a atenção nas divisões entre o individual e o coletivo, o material e o mental, a biografia e a história e a práxis e a teoria. Diferentemente da teoria originária, caracterizada por seu foco mais descritivo no desenvolvimento da personalidade e no uso da atividade como um princípio explicativo no nível das ações humanas, as versões ocidentais desenvolveram-se como tópicos interdisciplinares de investigação e de intervenção.

Assim, a teoria da atividade oferece uma abordagem ampla em sua essência, direcionada às dinâmicas desenvolvidas na cultura, na coletividade e em atividades socialmente distribuídas. Contudo, é importante destacar que a CHAT é uma teoria em evolução, aplicada em diversas áreas do trabalho e da aprendizagem organizacional, e, como tal, pode ser adaptada e desenvolvida em múltiplas disciplinas, variações e diferentes vertentes. Logo, é preciso definir que a vertente da CHAT a ser aqui aplicada é nomeada de “escola de Helsinque”, que tem Yrjö Engeström como seu principal expoente⁴⁴. Esta versão, bem influente no ocidente, toma por base os fundamentos essenciais da CHAT e desenvolve uma abordagem com foco nos sistemas de atividade do trabalho e nas práticas organizacionais, destaca-se principalmente pela modelagem dos sistemas de atividade como unidades principais de análise, por sua ênfase na orientação a objetos, no caráter da atividade dirigido por contradições e pelo uso da teoria da aprendizagem expansiva como instrumento de intervenção e transformação da atividade.

Para a exploração da CHAT neste estudo é necessário relacionar, de modo introdutório, alguns conceitos fundamentais, a saber: a) mediação por artefatos; b) diferença entre atividade e ação; c) objeto da atividade; d) abordagem dialética – relações internas, movimento, contradições e desenvolvimento; e) intervenção; e, f) unidade de análise. Estes fundamentos básicos são aplicados no decorrer do capítulo.

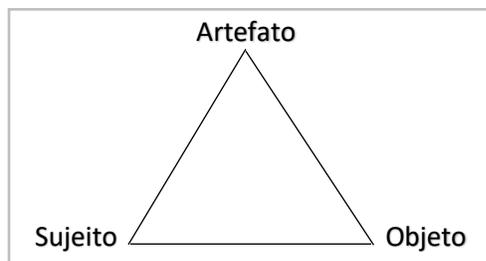
⁴⁴ A versão tem por base o livro seminal de 1987: *Learning by expanding. An activity-theoretical approach to developmental research*. Aqui citado em sua segunda edição de 2014 (Engeström, 2014).

a) **Mediação por artefatos** - o filósofo alemão Georg Hegel, foi o primeiro a chamar a atenção para o papel da atividade material e produtiva dos instrumentos de trabalho no desenvolvimento do conhecimento, onde o acúmulo de conhecimento pela sociedade ao longo da história, forma e influencia a consciência humana, objetivando a criação de artefatos pela humanidade (Engeström, 2014; Sannino, 2011). Para Cole (1996) artefato se refere a um aspecto do mundo material e conceitual, modificado ao longo da história da sua constituição através de ações humanas. Assim, os atores humanos usam e interagem com artefatos que evoluem culturalmente (Pelaprat & Cole, 2011), ou seja, estas ferramentas são mediadoras culturais usadas para modificar o mundo externo.

Vygotsky (1978) amplia o entendimento de mediação, originário em Hegel, indicando que a ação humana vai além do estímulo material, sendo também intermediada por “ferramentas psicológicas”. O autor emprega os termos ferramentas, instrumentos e artefatos como sinônimos de mediadores da atividade humana, propondo que estes sejam entendidos com recurso à ideia dos instrumentos físicos e mentais. A mediação humana por artefatos sugere que o indivíduo não seja analisado sem seus meios culturais, nem a sociedade sem a ação dos indivíduos em interação com os artefatos (Sannino, 2011). Assim, a mediação humana aponta para que a relação entre o sujeito e o objeto da atividade seja mediada por ferramentas ou sinais concebidos na sociedade e na cultura. A figura abaixo representa a ideia da mediação humana originada em Vygotsky.

Figura 27

Triângulo da mediação humana



Nota. Adaptado de Vygotsky (1978).

b) **Diferença entre ação e atividade** - Leont'ev (1981) acrescenta ao contexto da mediação que, não apenas no indivíduo, mas entre os seres humanos, as atividades evoluem mediadas por artefatos definidos no curso da história. Este acréscimo direciona para a

coletividade e para as interrelações globais, diferencia a ação individual da atividade coletiva, destacando o objeto e objetivo da atividade. No desenvolvimento desta concepção, a ideia é que as atividades derivam das ações, ou seja, um conjunto de ações podem formar uma atividade, de modo que as atividades não são entendidas sem ações e nem as ações podem ser compreendidas sem a atividade. Portanto, para perceber uma ação individual é necessário contextualizá-la na atividade à qual pertence, ocorrendo reciprocidade. Logo, ação e atividade são fenômenos distintos, mas unificados (Ilyenkov, 1982). No entanto, a diferença entre a ação individual e a atividade coletiva se mostra no fato de que as ações têm um objetivo imediato e localizado, mas o motivo real da ação está em algo mais amplo, o que geralmente é uma atividade coletiva com um objeto comum a ser alcançado.

c) Objeto da atividade - o objeto desempenha um papel central na CHAT. Neste sentido, Leont'ev (1981) argumenta que não há atividade sem um objeto, e que a razão para existência do objeto está relacionada a uma necessidade presente na sociedade. Assim, toda atividade tem uma necessidade social na qual os seres humanos perseguem, reproduzem e buscam transformar esse objeto por meio de ações, de modo a satisfazer tal demanda. O princípio básico do objeto é que este só pode ser alcançado coletivamente. Nessa perspectiva, o objeto é uma construção cultural e coletiva que possui uma vida histórica e cria sentido para atividade (Kaptelinin, 2005).

d) Abordagem dialética - a base conceitual da CHAT está fundamentada sob os princípios gerais do materialismo histórico e da dialética materialista⁴⁵. Isto tem por base a percepção de que as ações humanas são mediadas pelas ferramentas culturais (Sannino & Engeström, 2018). Tal abordagem é particularmente útil para a compreensão do sistema de relações, de modo a contribuir para acompanhar e analisar o movimento de transformação, aqui especialmente necessário para adaptação da atividade da RDM ao ambiente. Tolman (1981) considera que a compreensão da dialética pressupõe a análise de quatro elementos: 1) relações internas, 2) movimento próprio, 3) contradição interna; e, 4) desenvolvimento.

⁴⁵ Segundo Löwy (2010), muitos autores usam indistintamente os termos dialética materialista, materialismo dialético, filosofia da práxis, materialismo cultural, método dialético, materialismo histórico etc., para expressar a tentativa formulada por Marx para entender ou tentar enfrentar a relação entre as ideias e a ação, ou prática social. Na CHAT, a dialética é percebida como uma unidade de opostos que requer interação das forças contrárias para que algo novo seja desenvolvido.

1) Relações internas: é a ideia de que a relação entre um organismo e seu ambiente se dá de modo interno, no sentido de que o ambiente e o organismo se definem e se moldam mutuamente, em função das interações que ocorrem dentro de seu mundo de vida e da história. Assim, os elementos da atividade são dialéticos e não existem de forma independente. É a base para a importância da percepção da atividade de modo sistêmico.

2) Movimento próprio: a dialética estuda os fenômenos em seu auto movimento. Concentrado nas relações internas, podem-se explicar os movimentos por meio dos impactos causais que um tem sobre o outro. Para Tolman (1981), as coisas estão constantemente mudando e evoluindo, elas não existem sem movimento, sugerindo que a dialética vai além do movimento externo ou mecânico e que contemplar a existência de um fenômeno não permite compreender o começo e o fim de sua existência. Nesse sentido, a forma como se faz algo está sempre em movimento, ou seja, a atividade está sempre em movimento. Os autores na CHAT usam o movimento, ao menos, com dois significados. O primeiro para referir a compreensão da essência das coisas (Tolman, 1981) e o segundo com o objetivo da mudança da atividade para um estado futuro e melhor (Marshak, 1993). Tal movimento pode ser explicado através de contradições nas relações internas que constituem a atividade.

3) Contradições internas: a dialética materialista faz uso das contradições como a raiz de todo movimento, ao apontar que é apenas na medida em que algo tem uma contradição é que se move, tem um desejo e atividade (Marx & Engels, 2015). Uma contradição é uma tensão dinâmica historicamente acumulada entre forças opostas (Ilyenkov 2008), reveladas em distúrbios, muitas vezes tomando a forma de manifestações discursivas como dilemas e conflitos (Engeström & Sannino, 2011). Assim, conflitos, dilemas, distúrbios e perturbações podem ser analisados como manifestações das contradições (Sannino, 2005).

Realizar uma análise dialética das contradições, que emergem das interações de diferentes elementos e níveis, é necessário para fazer jus ao poder explicativo da CHAT (Roth & Lee, 2007). Isto porque o contraditório, apesar de poder ser visto como negativo, funciona como propulsor de evolução, movimento e transformação. Assim, as contradições tornam-se a principal fonte de transformação da atividade, é o princípio de seu auto movimento e a forma em que o desenvolvimento é moldado (Ilyenkov, 2008).

4) Desenvolvimento e transformação: a necessidade de desenvolvimento se deve quando a atividade não funciona de maneira suficientemente adequada, ou ainda não se tem o total conhecimento do que se deve fazer para melhorar o que existe no ambiente. Nas palavras de Tolman (1981), o que leva ao desenvolvimento de algo é o negativo que possui dentro de si mesmo, isso é que constitui a verdadeira dialética. O desenvolvimento, que sob o ponto de vista da dialética é a forma mais elevada de movimento, não significa o mesmo que crescimento – que carrega consigo o sentido de mudança quantitativa, mas refere-se a uma mudança qualitativa (Querol et al., 2014) que possibilita resultar na transformação da atividade. Portanto, o desenvolvimento permite ser entendido como um processo para resolver contradições, de modo que novos estágios qualitativos e formas de atividade possam emergir como soluções destas contradições.

Por sua vez, o conceito de transformação deriva do entendimento de Marx (1976) que considera o indivíduo um agente criador. Assim como desenvolvimento não é o mesmo que crescimento, para os teóricos da CHAT a transformação não é o mesmo que mudança. A mudança está ligada a tornar as coisas diferentes objetivando a melhoria, já a transformação resulta em algo novo e totalmente distinto, uma nova gama de possibilidades que não existia anteriormente, isto é, gera modificações significativas e fundamentais em diversas áreas de desempenho. Assim, é possível entender que o desenvolvimento pode resultar em uma mudança ou transformação da atividade.

Como neste estudo busca-se também a adaptação da RDM ao ambiente da investigação e como a expressão “desenvolvimento” pode causar certa confusão de entendimento por sua variedade de aplicações, o termo “transformação” será adotado como sinônimo de desenvolvimento que tem sua base na mudança. Isto porque a mudança ou transformação, conforme aqui expostos, atendem à necessidade adaptativa da RDM ao ambiente da investigação no contexto da universidade.

A abordagem dialética aqui apresentada é utilizada em outro aspecto fundamental na CHAT, a intervenção na atividade. Nesse sentido, a CHAT não se resume à análise e à compreensão da atividade, mas propõe uma abordagem de caráter intervencionista.

e) Intervenção - a base intervencionista da CHAT recorre a Marx, Vygotsky e Leont'ev, para justificar o uso da intervenção desde as suas raízes, argumentando que tais autores se envolveram em vários tipos de intervenções em ambientes diversos, como os experimentos transformadores (Bronfenbrenner, 1977) ou experimentos de modelagem genética (Zuckerman, 2012). A intervenção na atividade está concebida na dialética e justificada no auto movimento da atividade. A ideia é concebida no fato de que como a atividade está sempre em movimento, as contradições já existentes, materializadas nas relações internas, fazem com que a atividade naturalmente caminhe para a transformação. A intervenção serve para direcionar e adequar de modo mais célere e pertinente tal caminho natural. Nessa direção, os agentes interventores utilizam de modo fundamental a ideia de unidade para realizar a análise da atividade.

f) Unidade de análise - na realização das intervenções com o uso da CHAT, é necessário primeiramente definir a unidade de análise. Esta definição tem origem no pensamento de Vygotsky que sugeriu a análise baseada em uma unidade e não em elementos, defendendo que a ação culturalmente mediada do indivíduo reúne o que é necessário para constituir uma unidade de análise (Zinchenko, 1985).

A unidade de análise foi expandida por Leont'ev (1981) sob a diferenciação entre ação individual e atividade coletiva. Nesse sentido, a atividade coletiva, na relação entre o indivíduo e a sociedade, é a unidade mínima de análise que contribui para explicar a coerência de um conjunto de ações individuais (Kuutti, 1996). Este entendimento soma-se à ideia da unidade como um sistema, uma combinação mínima de relações internas e o conjunto de opostos que manifesta as qualidades e a dinâmica do todo. Portanto, na CHAT um sistema de atividade, mediado por artefatos e orientado pelo objeto, é a unidade mínima a ser analisada (Engeström, 1999).

5.2. MODELO DO SISTEMA DE ATIVIDADE

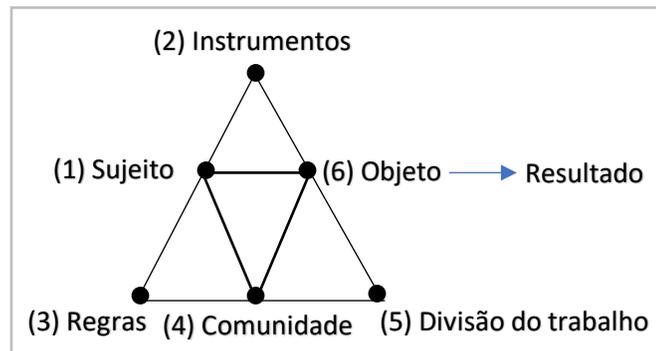
Engeström (2014) desenvolveu um modelo para o Sistema de Atividade (SA), o qual representa os relacionamentos básicos no sistema de mediação da atividade humana (Figura 27). O modelo proposto descreve os processos de mediação cultural – produção, distribuição e troca, os quais, de acordo com o autor, estão presentes em todas as atividades

coletivas que ocorrem em uma atividade. Um sistema de atividade pode ser entendido como uma formação coletiva (um grupo de pessoas em uma organização relativamente estável), que possui seus instrumentos (artefatos mediadores), regras, divisão do trabalho e a sua comunidade (mediadores organizacionais), orientada por um objeto partilhado. Nesse sentido, o modelo do SA amplia o triângulo individual de mediação humana de Vygotsky (1978), ao incorporar os mediadores sociais organizacionais. Esse processo colaborativo é idealizado através de seis componentes básicos.

Os seis componentes foram dispostos no modelo de um sistema de atividade (Figura 28) para funcionar de modo que na relação entre o (1) sujeito, os elementos mediadores e o objeto da atividade alcancem os resultados esperados. Os elementos mediadores do SA que podem ser identificados como recursos de mediação individual, são os (2) instrumentos – ferramentas ou sinais, as (3) regras sociais e institucionais que governam o sistema de atividade e a (4) comunidade de pessoas interessadas na atividade, o SA é organizado por uma (5) divisão do trabalho que serve para o alcance do resultado, (6) objeto coletivo.

Figura 28

Modelo de um sistema de atividade



Nota. Adaptado de Engeström (2014).

No detalhamento dos componentes básicos do modelo de um SA, o sujeito se refere ao indivíduo ou subgrupo, cuja posição e ponto de vista são escolhidos como aspecto principal a ser observado, de modo a alcançar o resultado pretendido. O objeto é a matéria-prima ou espaço de problema para o qual a atividade é direcionada. Os instrumentos são as ferramentas e sinais que contribuem para transformar o objeto em resultados. A comunidade

compreende os indivíduos e subgrupos que partilham o mesmo objeto. A divisão do trabalho refere-se à divisão horizontal de tarefas e a divisão vertical de poder e *status* da atividade. As regras são os regulamentos, normas, explícitas e implícitas, convenções e padrões que restringem as ações dentro de um sistema de atividade. Tais elementos estão sendo constantemente construídos e renovados, ou seja, o SA está em movimento.

O modelo do SA permite dar atenção aos componentes constituintes da atividade. No entanto, enquanto unidade de análise, o SA pode ser utilizado para identificar a relação e a formação dos elementos da atividade, revelar as contradições dentro e entre os sistemas de atividade que possuam o mesmo objeto e ainda modelar ou desenvolver um novo conceito da atividade para resolver as contradições internas, alterando assim os elementos do sistema para a transformação da atividade. Isto é possível de ocorrer porque os atores em articulação e através do SA, podem ser mediadores para uma dialética transformadora da estrutura a ser analisada. Tal dialética apresenta diversos atores a serem identificados, analisados e percebidos em suas complexidades. Assim, pode-se entender que o SA é uma maneira pela qual múltiplos olhares e vozes participam da atividade, o que indica a transformação como um processo colaborativo e dialógico no qual diferentes perspectivas se encontram, colidem e se fundem em múltiplos pontos de vista, tradições e interesses (Engeström, 2001).

Ao observar o modelo do SA, parece ser relativamente fácil “sistematizar” os atores, identificando quais se adequam as definições do modelo. No entanto, um SA é mais complexo do que a identificação ou soma mecânica dos seus componentes, é necessário ir além da sistematização inicial para detalhar como, através do SA, pode-se proporcionar a adequação a heterogeneidade do ambiente na universidade. Portanto, para que uma RDM no contexto da universidade, possa a partir da constituição do modelo do SA ser adaptada e assim adequar-se ao ambiente específico da investigação, de modo a proporcionar uma mudança qualitativa ou transformação da atividade, é exigido a expansão do objeto. Para tal, a CHAT faz uso de um tipo específico de aprendizagem, a aprendizagem expansiva.

5.3. A APRENDIZAGEM EXPANSIVA

A expansão na CHAT apresenta a ideia de que o SA, como está em movimento, transcende o contexto em busca de uma reorganização. Para conduzir tal expansão, Engeström (2014) sugere o uso da Aprendizagem Expansiva (AE), um tipo específico de aprendizagem que está ligada em aprender para transformar a própria atividade. O movimento expansivo nesse caso significa a transformação do objeto da atividade para um estado futuro e melhor (Marshak, 1993). Como já referenciado, o estado futuro e melhor, ou transformado, pretende ser neste estudo uma RDM adaptada ao ambiente.

Para explicar a aprendizagem expansiva, parte-se da ideia que na aprendizagem convencional o conhecimento a ser assimilado já existe, portanto sabe-se qual é a demanda e a solução a ser proposta. Na AE os sujeitos aprendem o que ainda não existe, ou seja, a AE parte do pressuposto de que os indivíduos e a coletividade aprendem continuamente algo que não é estável, nem mesmo definido ou compreendido com antecedência. Assim, a aplicação da AE é considerada particularmente útil em ambientes não tradicionais, híbridos e multiorganizacionais, onde a atividade necessita ser redefinida para produzir formas e padrões novos, ou onde estão a experimentar mudanças ou deparar-se com um problema novo e complexo, ou ainda em situações que ainda não se saiba qual é o problema a ser enfrentado (Engeström & Sannino, 2010).

Tendo por base a intervenção e através de um processo bem estabelecido, a AE envolve a reconstrução dos elementos do SA para expandir o seu objeto e conceito. O processo requer cooperação entre os sujeitos envolvidos na transformação da atividade coletiva. Isto propõe que a aprendizagem seja entendida como uma criação colaborativa multivocal de novos modelos e padrões para a atividade, mais amplos em termos de escopo, participação, sustentabilidade e complexidade do que o modelo existente, desenvolvida em um processo multifacetado de debate, negociação e orquestração, que envolve ações colaborativas para implementar e estabilizar o novo formato, resultando na transformação do objeto da atividade (Engeström, 2004; Engeström & Sannino, 2010).

As visões e iniciativas do processo da AE são voltadas para o futuro, isto é, ao ser confrontada pelas adversidades, a atividade coletiva enfrenta algum tipo de problema

transformacional, o que exige dos sujeitos da atividade uma mudança para aprender o que ainda não existe. Assim, a expansão do objeto pede um modo de compreensão das contradições internas do SA, e encontrar possibilidades para continuar a desenvolvê-lo, o que é possível através da análise de sua formação histórica, das contradições existentes no ambiente organizacional e da forma de superação de tais contradições.

As contradições como fonte principal da expansão do objeto funcionam como elemento impulsivo para a transformação da atividade. A AE visa identificar e superar uma contradição que leva a atividade a uma situação de crise. Assim, com a contradição ultrapassada, o objetivo transformativo pode ser realizado. Isto é possível tendo em vista que um SA está constantemente em movimento, sendo construído e renovado em consequência do desenvolvimento de tensões e contradições, dentro e entre os diferentes níveis e elementos do sistema. A partir dessas concepções, Engeström (2014) identifica quatro ordens de contradições, quais sejam: primária, secundária, terciária e quaternária.

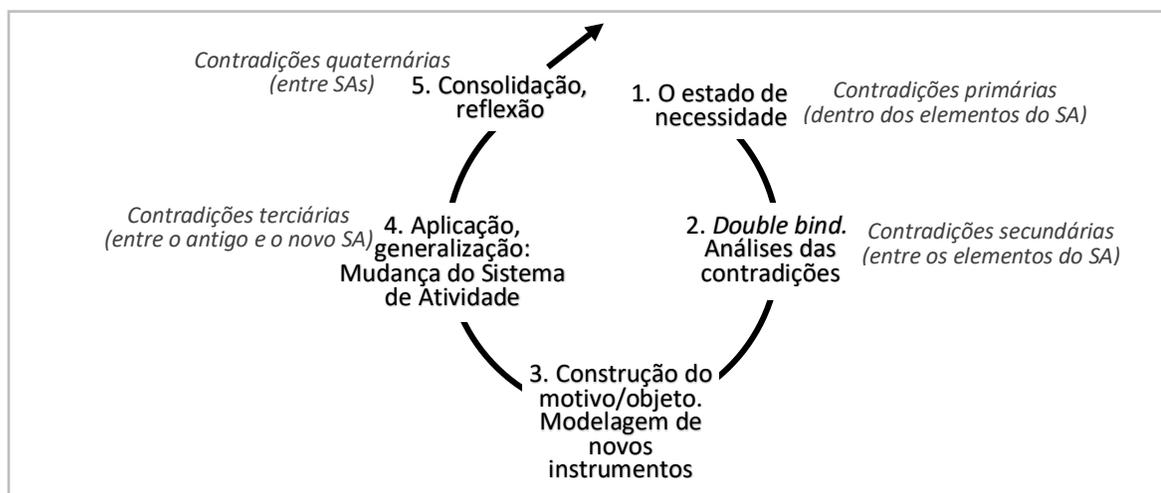
As contradições primárias podem aparecer dentro de cada um dos elementos do sistema de atividade, acontece entre o valor de uso e o valor de troca de um componente sistêmico. Uma contradição secundária manifesta-se abertamente entre dois ou mais componentes do sistema, acontece quando um SA adota um novo elemento de fora da atividade, ou quando algum elemento antigo colide com o novo. Contradições terciárias apresentam-se entre um modo de atividade recém-estabelecido e resquícios do modo de atividade anterior, gerando resistência e forçando o novo modelo a ser modificado. E por fim, as contradições quaternárias podem aparecer entre a atividade recém-reorganizada e seus SAs vizinhos, pois quando interagem, pode ser gerado tensões, perturbações e conflitos nas relações entre vários sistemas de atividades.

Identificado o problema originado em uma contradição, este pode ser resolvido com a transformação do sistema. Assim, na busca pela mudança, por algo novo, que visa superar tais problemas, Engeström (2014) propõe uma ferramenta para implementar a AE, a qual consiste numa representação gráfica de um ciclo, uma sequência de ações epistêmicas que ocorrem no aprendizado expansivo, nomeado de ciclo de transição expansiva (Figura 29). Tal ciclo é um processo de transição, para resolver contradições por meio de ações de aprendizagem no caminho da expansão do objeto. O processo envolve interações, quebra

de obstáculos que se movem para trás e para frente para chegar ao novo objeto. Contudo, a atividade resultante não necessariamente corresponde a um estado final, e a transformação pode continuar através de ciclos recorrentes de expansão. Ou seja, o modelo ideal do ciclo de AE é uma representação concebida para compreender as contradições internas do SA, captar a sua essência, encontrar possibilidades de resolução e continuar a desenvolvê-lo na transformação do objeto da atividade. O modelo associa os quatro níveis de contradição em uma sequência de cinco fases que ocorrem durante a aprendizagem expansiva.

Figura 29

Fases de um ciclo de transição expansiva



Nota. Adaptado de Engeström (2014).

Na primeira fase do ciclo, um questionamento pode caracterizar uma situação de insatisfação com a situação existente, um estado de crise ou um estado de necessidade urgente de fazer algo, no qual as contradições aparecem em sua forma básica, em geral correspondente a uma contradição primária. Durante essa fase, os sujeitos começam a discutir e a desafiar o propósito da sua atividade - seu objeto - e a forma atual de alcançá-lo. Há uma necessidade de mudança, mas ainda não há uma pressão urgente, pois é possível deixar a situação como ela se encontra (Engeström, 2014).

A segunda fase - *double-bind*⁴⁶ -, caracteriza-se quando uma contradição não pode ser resolvida apenas por meio de ações individuais, mas necessita de ações coletivas que

⁴⁶ Tem por significado uma situação momentânea de constrangimento que precede e sucede uma contradição.

possam levar para uma forma nova de atividade (Engeström, 2014). Tal contradição, começa a produzir desajustes e tensões entre os elementos do SA na perspectiva das contradições secundárias. Pessoas que participam da atividade sentem que não é mais possível continuar a fazer as coisas da maneira atual, mas elas ainda não sabem o que deve ser feito para resolver os problemas (Engeström & Sannino, 2012). Deve-se então analisar a situação para superação da contradição identificada e do problema apresentado.

Numa terceira fase, intitulada de “objeto”, inicia-se a construção do novo motivo da atividade, com a idealização da forma e da modelação de novos instrumentos. Nessa fase, a coletividade projeta uma nova atividade na qual o objeto é mais expandido, ou seja, mais amplo, já que inclui características mais desejáveis do que o anterior. No entanto, as mudanças podem ser simplesmente ajustes nos elementos do SA, como uma nova tecnologia ou maneira de fazer algo. Se a crise é grave, as pessoas podem desafiar todo o sistema, inclusive o propósito da atividade, redesenhando-a de uma forma mais expansiva.

A quarta fase, que corresponde à aplicação ou generalização, acontece quando a solução é modelada e a atividade idealizada pode ser implementada. Os instrumentos precursores da nova atividade são confrontados com a atividade antiga. A comunidade começa a materializar os planos de produção do objeto idealizado, aqui as tensões entre os elementos da nova atividade e os elementos da atividade anterior podem surgir, caracterizando as contradições terciárias. Esses desajustes podem ser causados tanto pelo desenvolvimento insuficiente dos novos elementos quanto por alguma incompatibilidade entre o novo e o velho (Engeström, 2014).

Uma vez que a nova atividade começa a tomar forma, segue uma quinta fase, a consolidação. Nesta etapa, a atividade é consolidada quando os novos conceitos são sistematicamente aplicados de uma forma repetitiva e explícita. A atividade mudada ou transformada se ajusta aos SAs vizinhos para se afirmar no contexto. Porém, antes de ser capaz de consolidar a nova atividade, é necessário resolver essas tensões com as atividades relacionadas que ainda seguem a velha lógica de produção, isto acontece resolvendo contradições quaternárias, pois a nova atividade tem que competir e se ajustar à dinâmica das atividades vizinhas (Engeström, 2014).

Na aplicação do modelo, uma transformação é realizada quando o objeto e o motivo da atividade são reconceituados, envolvendo um horizonte mais amplo de possibilidades do que no modo anterior. No entanto, o ciclo ou processo da AE não é uma fórmula universal de ações, fases ou estágios, provavelmente não seja possível encontrar um processo de aprendizagem coletiva que siga claramente o modelo típico-ideal.

Uma definição possível para o ciclo de transição expansiva, é que este seja um processo para resolver contradições, em evolução sucessiva, por meios de ações de aprendizagem, visando alcançar atividades futuras, ainda não definidas ou percebidas totalmente. Para tal finalidade, a CHAT faz uso de duas ferramentas básicas: o princípio dialético que segue a lógica de ascender do abstrato ao concreto, adaptado de Davydov (1988), e o conceito da Zona de Desenvolvimento Próximo (ZDP) proposta por Vygotsky (1978). Nesse sentido, a aprendizagem expansiva, enquanto ciclo de transição para alcance de atividades futuras, pretende ser uma jornada através do terreno desconhecido da zona de desenvolvimento proximal (Engeström, 2005). A AE utiliza o conceito da ZDP para delinear e compreender as tendências de expansão do objeto da atividade.

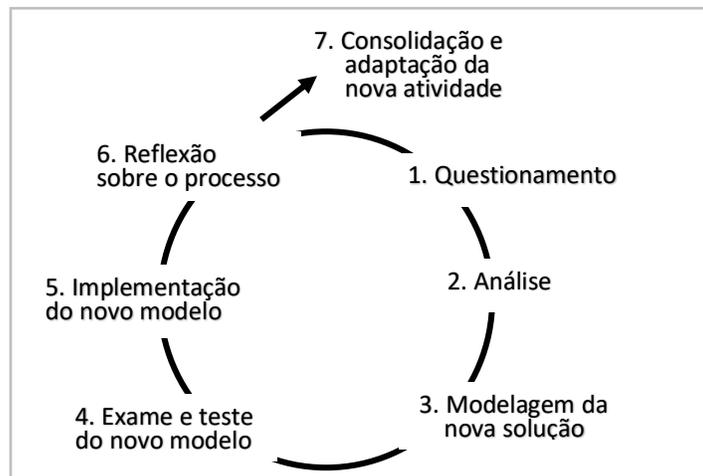
A ZDP, foi redefinida na CHAT para ser utilizada na dimensão coletiva como espaço de transição das ações para a atividade, ou seja, uma transição expansiva capaz de lidar com a aprendizagem e o desenvolvimento no nível das atividades futuras. Para isso usa-se tal “zona” como a distância entre o nível do desenvolvimento real e o nível do desenvolvimento potencial, determinado por meio da resolução de problemas. A ideia é que a distância entre o que pode ser alcançado no presente e o que pode ser adquirido no futuro, seja solucionado através da interação de pessoas e ferramentas culturais em uma transição para formação de um novo objeto, ou seja, a transformação.

Na mesma direção, visando a obtenção de uma nova solução transformadora, utiliza-se outro conceito basilar, o princípio dialético da ascensão do abstrato ao concreto. Adaptada das três ações de aprendizagem de Davydov (1988), o princípio pode ser aplicado de forma prática como uma sequência de ações epistêmicas, representado no modelo nomeado como espiral expansivo (Figura 30) (Engeström, 2001). Neste princípio, uma nova ideia ou conceito teórico é inicialmente produzido na forma da apreensão de uma abstração inicial, a abstração é enriquecida passo a passo e transformada em um sistema concreto. O

princípio é usado para apreender a essência de um objeto, traçar e reproduzir a lógica de seu desenvolvimento e de sua formação histórica em uma transição das ações individuais para uma atividade coletiva, qualitativamente nova, por meio da resolução de suas contradições.

Figura 30

Sequência de ações do espiral expansivo.



Nota. Adaptado de Engeström (2016).

As ações no espiral expansivo começam com os sujeitos individuais a questionar (1), criticar ou rejeitar aspectos práticos da situação atual da atividade e gradualmente se expande em um movimento coletivo. Ao analisar (2) coletivamente de onde vem o problema, causas ou mecanismos explicativos podem ser descobertos, permitindo modelar (3) uma solução ou uma “célula germinativa” em um meio observável e transmissível. Através do novo modelo pode-se examinar e testar (4) as ações, movendo-se para implementá-lo (5) por meio de aplicações práticas, enriquecimentos e extensões conceituais, e assim apreender sua dinâmica, potencialidades e limitações que permitam refletir e avaliar o processo (6) a fim de consolidar e generalizar (7) seus resultados expansivos em uma forma estável da nova prática. Este processo de AE, é caracterizado por uma discussão de múltiplas vozes, negociação e hibridização entre diferentes perspectivas.

Verifica-se que a aprendizagem expansiva utiliza diversas ferramentas e ações para o aprendizado em direção a transformação do objeto da atividade. Neste sentido, é necessário a percepção e colaboração entre as pessoas envolvidas no uso dos recursos disponíveis que combinados, podem cumprir o objetivo da AE. No entanto, também é

preciso organizar uma metodologia adequada para ser aplicada ao contexto em que o sistema de atividade precisa ser redefinido, objetivando produzir formas e padrões culturalmente novos para a atividade. No caso deste estudo, para atingir os objetivos declarados, tal organização é realizada através de uma proposta de ciclo metodológico.

5.4. UMA PROPOSTA DE MODELO DE CICLO METODOLÓGICO

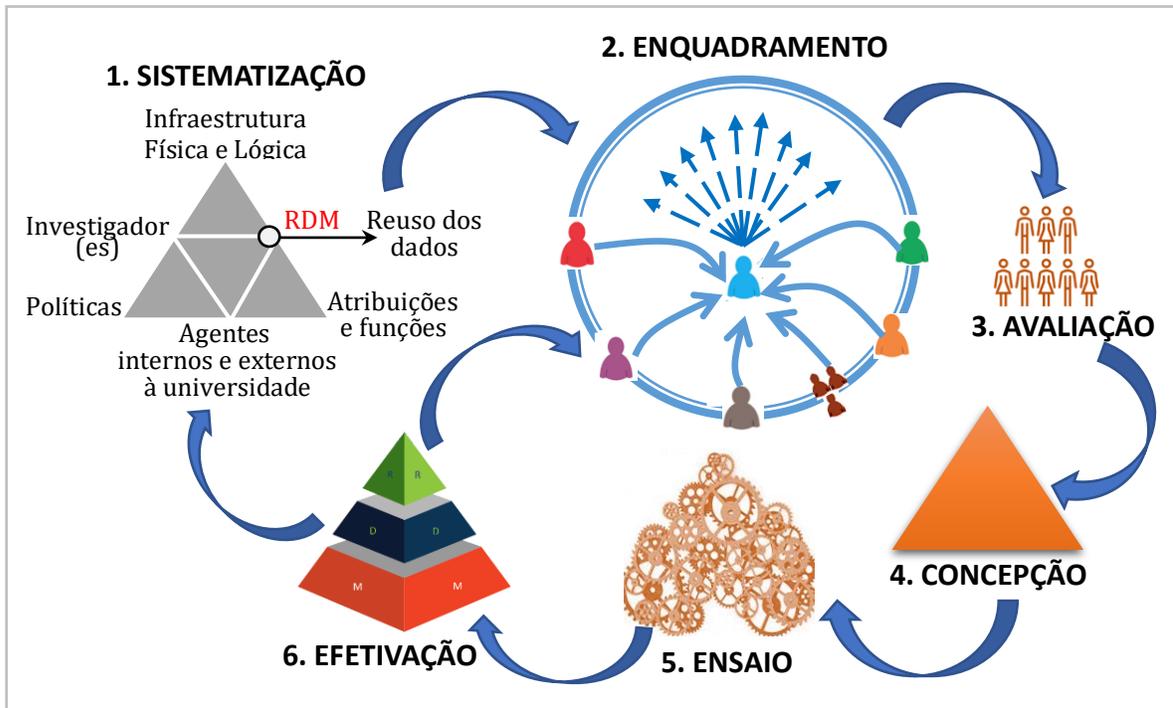
Apresenta-se aqui, em forma de modelo cíclico (Figura 31), uma proposta de metodologia para articulação dos atores e a adaptação da RDM ao ambiente complexo e heterogêneo da universidade. O modelo tem a finalidade de oferecer ferramentas para caracterizar o SA no ambiente, de forma que os relacionamentos dos atores possam ser analisados e as conclusões direcionadas para a expansão do objeto, resultando na transformação da atividade. A base para a proposta está fundada no construto apresentado anteriormente e no método intervencionista decorrente da CHAT, nomeado de Change Laboratory⁴⁷ - uma ferramenta teórico-metodológica para ampliar as possibilidades de investigação no campo da aprendizagem organizacional e gerar soluções para problemas complexos em diferentes setores (Querol et al., 2011; Virkkunen & Newnham, 2013).

Portanto, a finalidade do ciclo aqui proposto pode ser cumprida através de uma intervenção formativa – uma metodologia aplicada como instrumento de diálogo e de colaboração, onde busca-se que participantes-chave tomem iniciativa e sejam protagonistas na transformação da própria atividade. A título de exemplo, os elementos participativos no ambiente da RDM poderiam ser compostos por: a) um agente mediador que possa intervir para implementação e desenvolvimento do aparato teórico e metodológico; b) investigadores que trabalham diretamente nos projetos de investigação; e, c) profissionais envolvidos e/ou responsáveis pela RDM que podem pertencer aos quadros técnicos e gerenciais, associados aos projetos ou unidades de investigação. Assim, sob o construto realizado até agora, o modelo do ciclo metodológico é apresentado abaixo em seis fases, primeiro explicado de forma geral e depois detalhada cada uma das fases.

⁴⁷ O Change Laboratory foi implementado em instituições educacionais, hospitais e diversos locais de trabalho em cerca de 30 países, entre eles estão: Finlândia, Nova Zelândia; Brasil; Reino Unido; Estados Unidos; Japão; Espanha; África do Sul, Zimbawe e Lesoto (Querol et al., 2011; Virkkunen & Newnham, 2013).

Figura 31

Ciclo metodológico para a transformação e adaptação da RDM



O modelo primeiramente realiza a sistematização dos atores, de modo a produzir uma visão preliminar da RDM no ambiente e identificar os problemas vivenciados na atividade. Para tal, o ponto de partida é um delineamento do SA atual, juntamente com o levantamento das contradições, aqui também nomeado de modo simplificado de “problemas”. No segundo momento, é necessário comparar os problemas empíricos levantados com os problemas históricos da RDM, ou seja, as barreiras, conflitos e tensões no contexto, de modo a produzir um quadro analisável das contradições. Na terceira fase, para estabelecer as dimensões da transformação e identificar novas formas potenciais de organização, examina-se os problemas da atividade, identificando e propondo a superação das principais contradições encontradas. A quarta etapa é a idealização de forma qualitativa dos novos instrumentos. Assim, o novo modelo do SA é concebido na perspectiva da resolução das contradições analisadas. Na quinta fase, realiza-se o ensaio no ambiente, os testes no modelo idealizado. E na sexta fase, efetiva-se os novos instrumentos do SA. Caso novas contradições apareçam nesta última etapa, o ciclo recomeça a partir da segunda fase, para que possa cumprir novamente as etapas, até que as novas contradições sejam superadas.

Assim, a RDM por estas seis etapas pode ser transformada e adaptada de modo a gerar um novo SA no ambiente. Segue a explicação da finalidade de cada uma das fases constituintes da estrutura proposta, juntamente com um modo de uso e os conceitos da CHAT aqui trabalhados que fundamentaram as respectivas fases.

1) **Sistematização** – é a organização e visualização da articulação dos atores da RDM no ambiente. Esta ação tem a finalidade de obter o modelo do SA atual e possibilitar a análise das relações internas. A etapa proporciona, além da identificação e arranjo da articulação sistêmica dos atores na representação da RDM, um ponto de partida no processo deliberado de análise e transformação. Na dinâmica relacional entre os atores que compõem essa sistematização, têm-se que a RDM é o objeto comum, o sistema gere os dados de investigação através de uma articulação ideal, com a ajuda de mediadores individuais e coletivos, a fim de obter como resultado a reutilização dos dados de investigação.

A relação dos atores forma um sistema porque depende necessariamente de todos para que o objeto comum da atividade, a RDM, seja realizada corretamente. Assim, o **investigador** é o sujeito inicial escolhido para observação da sua perspectiva em relação aos mediadores individuais e coletivos do sistema, tanto o investigador, quanto os outros atores utilizam uma **infraestrutura física e lógica**, instrumentos para tratar e gerir os dados de forma adequada. No entanto, a concepção e uso de uma infraestrutura depende da divisão do trabalho, ou seja, de levar em consideração **atribuições e funções** bem definidas para o correto funcionamento da RDM. Seguindo no mesmo senso de articulação, aplicar as **políticas locais, nacionais ou internacionais** para a RDM, significa interferir no curso concatenado entre os elementos sistêmicos. E ainda, o relacionamento com as partes interessadas na RDM - **agentes internos e externos à universidade**, também gera consequências e influências recíprocas, sejam estas financeiras, políticas ou sociais, repercutindo em todos os componentes do SA para a gestão dos dados de investigação.

A utilização desta etapa passa pela identificação, arranjo e avaliação das relações dos atores da RDM no ambiente, possibilitando a indicação de um modelo idealmente articulado e o levantamento dos problemas que podem ocorrer sob tais articulações. Estes problemas são inadequações resultantes da historicidade do ambiente ou decorrente de situações específicas observadas de forma empírica. Nas observações, questões como o

status da atividade, dados históricos sobre eventos importantes, as práticas e o modo como a atividade é conduzida, os principais problemas enfrentados e os principais conceitos e ferramentas utilizadas na atividade podem ser levantados.

Em termos conceituais, a atividade é percebida como culturalmente mediada por artefatos, o que pode revelar especificidades de colaboração, interesses, comportamentos, questões técnicas, tradições ou hábitos que influenciam a adaptação às circunstâncias do ambiente e contribuem para a sua formatação. Tal abordagem liga dialeticamente os atores humanos ao contexto social, sem relegar a segundo plano a importância das atividades técnicas. O entendimento é que os elementos da atividade são dialéticos, sociotécnicos e não existem de forma independente. Isto evoca a percepção de que as influências, as motivações e os movimentos dos atores no envolvimento com as atividades da RDM são justapostos, possibilitando a análise da unidade, precisamente as dinâmicas relacionais que são importantes, ao mesmo tempo que as dinâmicas mesclam o material e o social.

Em termos práticos, ao identificar os atores e os arranjar de forma sistêmica, realiza-se uma análise das articulações, possibilitando a obtenção de uma visão preliminar da RDM atual no ambiente e um mapeamento dos problemas vivenciados na atividade.

2) Enquadramento – um quadro do SA para a RDM é aqui realizado com os problemas atuais e históricos levantados na etapa anterior e, organizados para posterior análise. O quadro deve tanto conter a visão geral do sistema, como os seus distúrbios e demais dados importantes obtidos na fase anterior. Nesta etapa, a ênfase está na organização das situações problemáticas verificadas na RDM, impeditivas que ela funcione de modo ideal para que os dados possam ser geridos e reutilizados. Por exemplo, as barreiras, tensões e conflitos mapeados no capítulo 3, podem servir de base para direcionar e selecionar os problemas observados no levantamento empírico. O enquadramento aqui proposto, contextualiza o caráter problemático associado à percepção do estado atual do sistema. Isto é necessário para analisar os distúrbios e suas origens históricas, em consequência possibilita a compreensão das contradições que afetam o sistema de atividade, pois são elas que motiva os indivíduos a realizarem as mudanças necessárias.

Durante esta fase, pode ser visto o chamado “estado de necessidade”, representado no ciclo de transição expansiva, onde uma situação de insatisfação com a circunstância atual, um estado de crise e a necessidade de mudar são estabelecidas. No mesmo sentido, a primeira ação do espiral expansivo é aqui aplicado em termos de questionamento da atividade, pois começa-se a discutir o motivo da atividade (o objeto) e as formas atuais de como realizá-la (as ferramentas), bem como os métodos e as tecnologias existentes.

Em síntese, nesta etapa são selecionados, organizados e discutidos de modo primário os problemas atuais e históricos do SA, sem ainda projetar soluções definitivas, mas de forma que o processo reflexivo possa ser desenvolvido mirando o futuro da atividade. O enquadramento deve organizar os elementos, relações ou eventos que são problemas históricos, juntamente com os problemas atuais que impedem o correto funcionamento da RDM. A dinâmica e objetivo desta etapa poderia ser inserida juntamente com a fase anterior (Sistematização) ou posterior do ciclo (Avaliação), porém dada a característica de dúvidas e questionamentos prévios que singularizam a RDM no ambiente, é necessário o realce através desta fase específica para a organização e discussão inicial exclusiva sobre os problemas, mas como qualquer modelo, pode ser adaptado a época e ao contexto a ser aplicado.

3) Avaliação – embora haja análise e avaliação em todas as fases do ciclo, esta etapa caracteriza-se por ser uma avaliação dos problemas direcionados à transformação, ou seja, a partir do quadro gerado na fase anterior, realiza-se uma reflexão aprofundada do passado e das ações no presente, para projetar um futuro ainda não instanciado. Assim, dois sentidos de problemas enquadrados anteriormente são avaliados nesta etapa, os problemas atuais e os históricos, ambos com o objetivo de determinar a origem contraditória da RDM no SA analisado. Seguindo este propósito, as causas ou mecanismos explicativos são descobertos, e então pode-se passar para uma situação de superação da contradição identificada, possibilitando a projeção futura do SA.

A avaliação dos problemas históricos fornece um panorama e uma hipótese acerca das contradições internas, o que favorece e direciona a exploração das práticas atuais realizadas na atividade. Assim, a análise dos problemas e das práticas produz uma imagem mais detalhada em relação às causas das perturbações encontradas, bem como do papel dos atores do SA em relação ao surgimento dos problemas. Mediante a análise nestas

perspectivas, a partir das contradições na estrutura do SA, pode-se obter uma representação concisa de suas causas sistêmicas, possibilitando superar os problemas superficiais e subjacentes da atividade presente.

Com base nos referenciais teóricos dos capítulos anteriores, a visualização de uma contradição que pode ocorrer nesta etapa é ilustrada através da seguinte inferência: um problema alegado pelos investigadores para a não partilha dos dados é a desconfiança quanto ao uso inadequado ou interpretação incorreta destes, porém é comum que a partilha seja uma exigência das agências que financiam a investigação. Assim nasce uma inadequação e gera um problema para a RDM, que neste caso é caracterizado como uma contradição primária. O resultado da análise desta contradição deve ser um novo princípio generativo, um método prático para superar a contradição geradora do problema dentro do sistema.

Ratifica-se que o exame dos problemas atuais e dos antecedentes históricos do SA, sob o propósito apresentado, é uma construção coletiva e participação colaborativa, onde os próprios sujeitos são agentes de transformação no processo, pois as contradições reveladas por esta avaliação define o conhecimento necessário que os participantes precisam obter para avançar em busca de soluções. Assim, é possível projetar a atividade futura, equacionando as contradições encontradas no presente.

Em termos conceituais, esta etapa corresponde a segunda fase do ciclo de transição expansiva, destacando o surgimento das duas primeiras ordens de contradição. Do mesmo modo, a etapa adequa-se na segunda ação do espiral expansivo. A conjugação destes fundamentos com as práticas para descobrir e superar contradições é a essência desta fase.

Resumidamente, a aplicação da etapa de avaliação é realizada para revelar as estruturas e contradições internas no SA, oriundas das situações problemáticas encontradas na atividade da RDM. Em consequência, a avaliação apresenta a estreita relação com uma futura modelagem do novo SA, pois para resolução das contradições deve-se criar germes de novas práticas e soluções inovadoras que apontam para as possibilidades da transformação da atividade. Para tal, a análise precisa estar ligada, retrospectivamente em perspectiva histórica acerca da origem dos problemas atuais, quanto prospectivamente, aos

desenvolvimentos futuros possíveis. Nesse sentido, esta fase possibilita uma avaliação para fornecer os dados e conceber um novo modelo para adaptação da RDM no ambiente.

4) Concepção – é a modelagem da (re)conceitualização da atividade no caminho para a transformação. Através desta modelagem de uma nova estrutura possível para a atividade, representa-se a perspectiva da solução das contradições observadas na etapa de avaliação. Define-se aqui a concepção qualitativa do novo modelo para a RDM, estabelecendo as dimensões da transformação na expansão do objeto, proporcionado pela formação de novos mediadores. Neste sentido, duas são as concepções a serem realizadas: 1) as visões futuras da atividade, onde os novos mediadores são instanciados, tendo por base não apenas uma classificação do que seja ou não desejável, mas a avaliação do SA, sob a perspectiva da resolução das contradições internas; e, 2) um modelo explícito e simplificado da nova ideia, uma célula germinativa, que explique a situação problemática e ofereça uma perspectiva para resolvê-la, através de novos mediadores que influenciará a atividade para sua transformação.

Nas duas concepções, que podem ser organizadas através de um plano de ação, são definidas novas formas e novas ferramentas com as quais os participantes começarão a realizar experimentalmente a visão de futuro. Por sua vez, a ação de modelar uma nova solução, materializa-se mediante a produção de versões e modificações das ideias originais, examinando-as até que se chegue a um modelo aceitável. Esta dinâmica encontra sua base teórica na terceira fase do ciclo de transição expansiva, juntamente com a terceira ação do espiral expansivo. Assim, a concepção de novos instrumentos é realizada pelos participantes na qual a RDM pode ser transformada. O modelo também deve permitir que a nova forma idealizada, possa ser legitimada através de testes sucessivos para posterior implementação.

5) Ensaios – nesta fase, o objetivo é levar as ideias abstratas levantadas e modeladas na etapa de concepção para um nível mais concreto e mais próximo de ser realizado. Assim, as mudanças modeladas podem ser testadas e/ou validadas na atividade real em um programa piloto que precede a efetivação do novo SA. Aqui, os participantes devem ser os protagonistas, e o diálogo interativo deve garantir uma perspectiva multivocal. Os ensaios do novo modelo ocorrem com o objetivo da exequibilidade da relação entre o novo sistema e o sistema atual da atividade. Estes direcionamentos têm o objetivo de antecipar o que pode auxiliar a transformação, bem como identificar os obstáculos que se

podem encontrar na atividade real. Parte da tarefa dos testes é continuar a concepção do novo modelo no nível dos detalhes, tendo em vista que, durante essa fase, novos problemas podem surgir na criação da nova forma da atividade. Cabe aqui avaliar se o fluxo segue para a implementação ou retorna para o nível anterior. Esta etapa está relacionada com a quarta ação do espiral expansivo e não há um correspondente claro no ciclo de transição expansiva.

6) Efetivação – são as ações para uso do novo modelo no ambiente, ou seja, os novos instrumentos, regras, modos de dividir o trabalho e de colaboração são postos em prática no SA. Nesse contexto, a efetivação envolve os conceitos de generalização e consolidação, ambos referenciados no ciclo de transição expansiva e nas ações do modelo do espiral expansivo. Portanto, a efetivação é um processo gradativo e adaptativo de generalização e consolidação, onde os resultados também devem ser avaliados. Logo, esta etapa pode ser observada sob tais características e o possível surgimento de novas contradições.

A efetivação refere-se ao uso do modelo em uma situação da realidade ampliada, não mais por ensaios em um “piloto”, mas posto em produção alargada no ambiente, é um processo de adaptação, enquanto os novos artefatos são repetitivamente utilizados de uma forma explícita. Aqui é preciso destacar que ainda cabe alteração nos mediadores para adaptação à nova realidade, pois na busca por espaço, o novo SA interage com as atividades vizinhas sob o mesmo objeto. Isso significa que os componentes do SA podem experimentar conflitos entre a antiga e a nova forma de fazer e pensar. Nessa perspectiva, a implementação do modelo novo e dos elementos novos da atividade podem conduzir para as contradições terciárias, onde todos os elementos do SA, enfrentam contradições entre sua forma existente e as aplicações do novo modelo.

Na atividade transformada, também emergem contradições entre esta e outras atividades na qual é dependente ou que dependam delas funcionalmente, são as contradições quaternárias. A resolução das contradições quaternárias leva a estabilidade da nova forma do sistema de atividade. Assim, a nova forma da atividade evolui para transformação e estabiliza-se com a resolução dessas contradições.

O modelo do ciclo metodológico aqui apresentado segue um fluxo que deve ser considerado flexível e didático em suas etapas, pois, em alguns casos, as fases podem acontecer também de forma transversal ou em sentido contrário. Assim, a proposta pretende ser aplicada de modo direcionado, mas também através de uma construção multivocal, proporcionar flexibilidade para que se perceba e acrescente outros métodos e fluxos pertinentes no desenvolvimento das fases e para a solução dos problemas. Nesse sentido, questões permanecem abertas, na qual o modelo não pretende ter todas as respostas, cabendo através da sua aplicação a possibilidade de resolução e expansão do próprio modelo para a transformação da atividade no ambiente.

5.5. CONSIDERAÇÕES SOBRE A ARTICULAÇÃO DOS ATORES

Este capítulo produziu um modelo de ciclo metodológico para conceber um sistema articulado dos atores para a RDM, bem como os parâmetros que podem permitir a sua adaptação ao ambiente heterogêneo e complexo da universidade. Tal resultado possibilita disponibilizar instrumentos estruturantes para que a RDM, suas articulações e práticas, bem como todo o ambiente que a permeia, possam ser compreendidos e organizados em ações coordenadas para uma partilha adequada dos dados de investigação. Assim, o estudo estabeleceu uma proposta para a articulação e adaptação contextual dos atores para a gerir e partilhar os dados, cumprindo o quarto objetivo da tese.

Para obter o resultado deste capítulo, foram exploradas as ferramentas metodológicas disponibilizadas pela teoria da atividade, primeiramente através da fundamentação que apresentou a base conceitual a ser utilizada, sendo depois conhecidos os componentes e os relacionamentos no modelo do SA, para a identificação dos atores e suas articulações. Em seguida, ressaltou-se a aplicação da aprendizagem expansiva como forma de transformação da atividade. Por fim, foi concebido o modelo de ciclo metodológico como produto do capítulo.

O modelo aqui concebido permite a identificação, arranjo e articulação sistêmica dos atores e viabiliza a adaptação da atividade da RDM ao contexto da universidade. Assim, foi percorrido o caminho para que as duas questões complementares apresentadas neste capítulo: como deve ser a articulação dos atores? e como a RDM pode adaptar-se ao ambiente complexo e heterogêneo da universidade? fossem respondidas. A primeira, na seguinte forma:

a articulação dos atores deve acontecer sob a perspectiva de um modelo sistêmico, de base sociotécnica, que relacione os atores em torno do seu objeto – a RDM –, mas também que seja orientado por uma necessidade social - a partilha e a reutilização dos dados –, considerando a atividade como culturalmente mediada por artefatos e sob constante movimento. Na segunda questão, a RDM pode ser adaptada e adequada as especificidades do contexto da universidade, através de ações cíclicas, que proporcionam o aprendizado para a análise e a transformação da RDM, tendo por base propulsora os problemas levantados no ambiente.

A abordagem do modelo de ciclo metodológico conforme apresentado se inscreve como parte integrante do modelo final para a RDM, pois esta abordagem proporciona uma estrutura necessária para que os dados de investigação sejam geridos e partilhados de forma adequada, tendo em vista a possibilidade de sistematização dos atores aqui definidos e de uma RDM adaptada ao ambiente, superando, em parte, o problema da complexidade e heterogeneidade do contexto. Embora o modelo deste capítulo possa proporcionar a sistematização e uma adaptação, também é possível ser ampliado para uma rede de sistemas de atividade, cuja aplicação pode ser gradativamente expandida, de forma a abranger todo ambiente da universidade. Cabe ressaltar que a aplicação deste modelo cíclico, sem observar de modo detalhado às práticas essenciais para a RDM, a partir do ambiente do investigador e da investigação, ainda limita a capacidade da RDM para uma partilha adequada dos dados objetivada nesta tese. Nestas considerações, foi pressuposto teoricamente que tal abordagem cíclica é compatível com a realidade a ser adaptada. No entanto, é preciso verificar se tal realidade, representada no campo empírico pelas Unidades de Investigação e Desenvolvimento da Universidade de Coimbra, realmente pode ser adaptada sob os atores aqui definidos e articulados no mesmo propósito da RDM. Assim, o próximo capítulo apresenta principalmente os resultados da teoria e os dados do campo empírico para análise e discussão, de modo a verificar esta e outras pressuposições, no confronto entre a abordagem teórica e a realidade, a partir do que foi desenvolvido até agora nesta investigação.

6. RESULTADOS E DISCUSSÃO

A partir do objetivo geral – propor um modelo para a RDM (*Research Data Management*) que possibilite uma partilha adequada dos dados no ambiente da universidade, tendo por base a articulação dos atores e o uso de práticas essenciais para a gestão –, este capítulo apresenta, analisa e discute os resultados provenientes da literatura e os dados obtidos no campo empírico da investigação. A apresentação dos resultados e sua discussão é estruturada conforme os objetivos da tese, sob os três eixos do estudo, nomeadamente: a partilha dos dados de investigação, a articulação dos atores e as práticas para a RDM. Assim, o capítulo tem início apresentando os resultados obtidos da literatura e o seu nexó teórico, para que nas seções seguintes sejam expostas as entrevistas e as respectivas discussões. Neste sentido, sempre que houve correspondência entre os dados relacionados ao roteiro das entrevistas e os resultados provenientes da análise da literatura, foram estes discutidos lado a lado.

6.1. ARTICULAÇÕES E PRÁTICAS PARA A PARTILHA DOS DADOS – NEXO TEÓRICO

O propósito principal do estudo da literatura pelos eixos teóricos foi fundamentar esta investigação, mas também dela se obter uma relação conceitual entre a articulação dos atores, as práticas essenciais para a RDM e a partilha adequada dos dados de investigação, objetivo específico cinco da tese. Assim, da leitura e discussão de obras relevantes acerca da ciência aberta, da teoria da atividade e de modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, foi possível obter os resultados a seguir apresentados para serem analisados e proporcionar o estabelecimento do nexó teórico objetivado, e, ao mesmo tempo fundamentar o modelo para a RDM na universidade.

A construção da relação conceitual que evidencia o nexó teórico, inicia-se aqui apresentando os resultados obtidos para caracterizar uma partilha adequada dos dados de investigação, fundamentado pelos princípios e práticas da ciência aberta no contexto da partilha dos dados em acesso aberto, apresentados ao longo do capítulo 3. Tais resultados, sintetizados, produziram primeiramente uma proposta de organização relacional (Figura 5), constituída pelos princípios e práticas da ciência aberta agrupados em aspectos temáticos que, se relacionados e utilizados de modo lógico e intencional possibilitam a partilha dos

dados de investigação em acesso aberto. Nesse sentido, tal organização relacional temática é representada pelos seguintes aspectos (Tabela 10): 1) científicos; 2) colaborativos; 3) legais; 4) *openness* (abertos); 5) pessoais; 6) sociais; 7) sustentáveis; e, 8) técnicos e materiais.

Os princípios e práticas da ciência aberta também apontaram comportamentos geradores de tensões e conflitos, principalmente sob a questão do acesso aberto. As categorias geradoras desses comportamentos (Tabela 11), podem influenciar na partilha dos dados de investigação em acesso aberto. Tais influências foram sintetizadas e especificadas como sendo de ordem: a) individual; b) institucional e c) técnica. São questões cognitivas e sociotécnicas, permeadas por fatores interrelacionais que podem influenciar na ação da partilha (Figura 6). Os atores a seguir foram referenciados como de importante influência no ambiente: a) agências de fomento; b) editoras científicas; c) bibliotecários; d) informáticos; e) Universidades e seus gestores; e, f) as diferenças das áreas disciplinares.

Ainda sobre a geração de tensões e conflitos, foram coletadas as seguintes preocupações e dificuldades que podem influenciar o investigador na ação da partilha: a) a falta de conhecimento ou sensibilização; b) as perdas de oportunidade; c) o tratamento com as informações sigilosas; d) o esforço adicional; e) a desconfiança com o plágio; f) a falta de usabilidade das ferramentas; g) as exigências das políticas mandatórias; h) o receio de desrespeito com os acordos de *copyright*; i) a falta de clareza quanto aos direitos do autor; j) a falta de diálogo e promoção do acesso aberto; l) a falta de interação e apoio técnico; m) a falta de políticas claras; e, n) a cultura disciplinar e a tradição de não partilhar.

O ambiente da investigação e os elementos gerais e basilares para a abertura dos dados de investigação (Figura 7) também foram observados, através do contexto que conta com os seguintes elementos a serem levados em consideração para que os dados sejam partilhados em acesso aberto: a) a comunicação da ciência; b) as políticas e mandatos; c) os metadados; d) os princípios FAIR (*Findable, Accessible, Interoperable and Reusable*); e, g) os repositórios em acesso aberto. Todo este conjunto, é sustentado pelos princípios e práticas da ciência aberta orientados pelos objetivos do acesso e dos dados abertos.

A partir das seleções anteriormente apresentadas, a seção final do capítulo 3 sintetiza e expõe através de treze itens, os elementos a serem considerados para uma partilha

adequada dos dados de investigação no ambiente da universidade, quais sejam (Tabela 14): 1) o acesso aberto; 2) o investigador na investigação local; 3) a ação da partilha; 4) as formas e os meios da partilha; 5) os princípios e as práticas da ciência aberta; 6) a organização relacional por aspectos temáticos; 7) as tensões e conflitos; 8) as barreiras; 9) os benefícios; 10) as especificidades do contexto; 11) as diferenças disciplinares; 12) os agentes do sistema de comunicação científica; e, 13) os atores para a abertura dos dados de investigação.

Considerando os treze itens acima, também foram destacados os elementos contextuais a serem considerados para uma partilha adequada dos dados de investigação na universidade (Figura 8), assim apresentados: a) as influências; b) as especificidades e os agentes relacionais; c) o movimento realizado para a partilha dos dados; d) as infraestruturas digitais em acesso aberto; e, e) a reutilização dos dados pelos investigadores e pela sociedade.

E por fim, os resultados obtidos no capítulo, também mostraram, além de quais elementos devem ser considerados, como deve ser realizada uma partilha adequada dos dados de investigação, da seguinte forma: 1) centrada no investigador; 2) de modo sociotécnico; 3) considerando o aspecto cognitivo do investigador; 4) fundamentada nos princípios e práticas da ciência aberta para a abertura dos dados e partilha em acesso aberto; 5) organizada; 6) flexível em relação ao ambiente; 7) sistêmica em relação aos atores; e, 8) direcionada para a reutilização dos dados.

A partir do arcabouço apresentado, uma partilha adequada dos dados pode ser definida como: um processo sistêmico e organizado para a abertura dos dados de investigação, realizada por um conjunto de atores associados, centrada no investigador como ator principal, sob o qual devem ser observadas as influências cognitivas, contextuais e sociotécnicas. Fundamentada no uso relacional e intencional dos princípios e práticas da ciência aberta, o processo para a partilha dos dados deve ser suficientemente flexível para abarcar as especificidades das áreas de conhecimento, as orientações e as exigências das políticas, bem como as documentações que facilitam o uso atual e futuro dos dados, sendo assegurada por infraestruturas digitais e sempre que possível na forma do acesso aberto, com todo o processo objetivado para a reutilização dos dados de investigação.

Do construto teórico que proporcionou a caracterização acima, destaca-se aqui o fato de que a comunidade científica internacional empreende esforços para estabelecer diretrizes e parâmetros para a partilha dos dados de investigação em acesso aberto (Borgman, 2012) e, ao mesmo tempo, empenha-se para remover as tensões e conflitos que cercam a temática da abertura dos dados, normalmente causado pela condição do acesso aberto (Ferguson et al., 2014; Rowley et al., 2017). Destas ênfases, é possível pressupor que tais esforços, centrados no movimento a favor da ciência aberta e tendo por base o acesso e os dados abertos, são movidos por variados interesses e motivações necessárias que provocam a comunidade da investigação em âmbito mundial (Bezuidenhout et al., 2017; UNESCO, 2021). Assim, caracterizar uma partilha adequada dos dados de investigação, condensa em si várias expectativas conceituais observadas na literatura, principalmente nos direcionamentos da ciência aberta e no contexto do acesso e dos dados abertos.

Ainda que a importância da partilha dos dados em acesso aberto seja reconhecida na ciência aberta e enfatizada no meio científico, é no uso de ações práticas realizadas pelos investigadores no processo de investigação, que a partilha encontra sua forma ideal a ser considerada na RDM (Pasquetto, Randles & Borgman, 2017). Ou seja, uma partilha adequada dos dados de investigação encontra sua base na ciência aberta, mas necessariamente também está associada a práticas para a RDM no processo de investigação.

As práticas consideradas como essenciais para a RDM, em uma partilha adequada dos dados, constitui o segundo eixo do estudo para permitir a concepção do nexos teórico. Nesse sentido, os resultados do capítulo 4, provenientes do estudo de nove modelos de ciclos de vida, que representam os fluxos da RDM, da curadoria e dos dados de investigação, formaram a base conceitual para a identificação das práticas essenciais para a RDM que possibilita uma partilha adequada dos dados. Primeiramente foi construída uma visão geral sobre as diferenciações, convergências e limites entre as práticas de curadoria digital e da RDM no processo de investigação, considerando também o ciclo de vida dos dados (Figura 9), no contexto da universidade. Isto foi necessário para que os modelos fossem melhor analisados, mas também para definir no resultado das práticas essenciais, o que concerne às práticas para a realização da RDM e o que se refere às práticas específicas de curadoria digital.

As diferenciações, convergências e limites entre os elementos no contexto, resultaram nas seguintes afirmativas: a) as práticas para a RDM abrangem a curadoria, o processo de investigação e o ciclo de vida dos dados; b) as práticas para a RDM remetem às ações que precedem e excedem o processo de investigação; c) a diferenciação entre a RDM e a curadoria está nos relacionamentos internos e externos à universidade que ultrapassam os limites das ações para tratamento direto com os dados; d) o processo de investigação contém as práticas de curadoria digital e o ciclo de vida dos dados; e) as práticas de curadoria fazem parte, mas transcendem o processo de investigação em ações específicas; f) o ciclo de vida dos dados contém as práticas de curadoria relacionadas à RDM e estão contidos, mas não limitados, ao processo de investigação; e, g) todas as práticas têm como sustentação e interseção os princípios da ciência aberta e o acesso aberto. Para identificar as diferenciações, convergências e os seus limites, também é necessário atentar e associar as práticas para a RDM e da curadoria ao ambiente específico da universidade.

Os princípios da ciência aberta e o acesso aberto são os fatores que unem todas as práticas referenciadas. Tais práticas precisam ser observadas e aplicadas com clareza no ambiente da RDM na universidade, de modo a aumentar a possibilidade de uma partilha adequada dos dados de investigação. Os parâmetros que conceberam as práticas nomeadas de essenciais a serem observadas para a RDM (Figura 25), foram constituídas por dois tipos distintos de práticas, identificadas como: 1) sequenciais e 2) contextuais. As práticas sequenciais (Figura 24) são representadas por etapas realizadas pelo investigador, apresentada através do seguinte fluxo de ações: a) planejar; b) recolher; c) analisar; d) partilhar; e, e) reutilizar. As práticas contextuais, resultado da análise da tabela 19, que se caracteriza por abranger duas ou mais etapas sequenciais, foram apresentadas nas seguintes ações: a) descrever; b) proteger; c) estruturar; d) monitorar; e, e) preservar. As práticas sequenciais e contextuais, envolvidas e interrelacionadas, formam as práticas essenciais.

No entanto, de modo a evitar que este elenco de práticas essenciais para a RDM tornem-se um tanto quanto confusas e imbricadas (Figura 26), possibilitando uma justaposição de trabalho, somado à necessidade de distinção entre as práticas para a RDM e de curadoria no ambiente, faz-se necessário, ratificar que tais práticas essenciais sejam associadas ao ambiente e ao conhecimento do contexto em que estão sendo realizadas.

Os resultados acima apresentados direcionaram para que as práticas essenciais a serem utilizadas para a RDM, em uma partilha adequada dos dados, no ambiente da universidade, pudessem ser definidas como: um conjunto de ações para gerir e partilhar os dados de investigação de forma adequada, sendo constituído por um fluxo de etapas sequenciais iniciadas com o planeamento e realizadas através da recolha e análise dos dados, tendo a sua finalização com o ato da partilha. O fluxo é permeado por práticas contextuais que abrangem duas ou mais etapas sequenciais, de modo a garantir a descrição e a monitoração dos dados, mas também a constituir a proteção legal e técnica necessária, bem como os recursos estruturais para preservação, gestão e a partilha dos dados de investigação. Sendo direcionadas ao acesso aberto, sob os princípios da ciência aberta e objetivada para a reutilização dos dados, as práticas essenciais devem ser associadas ao ambiente e ao conhecimento contextual, proporcionando também a distinção entre as práticas da RDM e da curadoria digital, de modo a evitar a justaposição de trabalhos.

É possível perceber que a base teórica para a realização das práticas essenciais encontradas para a RDM, a serem utilizadas em uma partilha adequada, também se encontram sob o acesso aberto e nos princípios e práticas da ciência aberta. É neste ponto que as práticas essenciais e a partilha adequada dos dados de investigação têm a sua junção estabelecida. Ou seja, ambas recorrem de modo fundamental ao acesso aberto e à ciência aberta para realização e direcionamento das suas atividades. Tal ponto de ligação pode ser visto quando os elementos teóricos, constantes nas definições aqui propostas sobre as práticas essenciais para a RDM e em relação a partilha adequada dos dados são destacados nesse sentido. Ou seja, pode-se verificar que uma partilha adequada dos dados de investigação é observada sob “um processo sistêmico e organizado para a abertura dos dados de investigação”, “assegurada (...) sempre que possível na forma do acesso aberto” e “fundamentada no uso relacional e intencional de princípios e práticas da ciência aberta”. Por sua vez, é fundamental verificar que todas as práticas essenciais para a RDM “estão sustentadas sob os princípios da ciência aberta e direcionadas ao acesso aberto”.

As práticas para a partilha dos dados de investigação, por serem diversas e abrangentes, devem estar bem fundamentadas para poderem ser utilizadas de modo adequado. É neste sentido que se percebe a importância da ciência aberta e do acesso aberto

para essa construção, pois embora sejam diferentes (Jorge & Albagli, 2017), a relação entre a abertura, a ciência aberta e o acesso aberto estão conceituados de forma interdependente (Willinsky, 2005), de modo a estarem interligados no uso de práticas para que a partilha dos dados aconteça (Banović, 2021; BOAI, 2002; David, 2008; Gezelter, 2011). Assim, tanto a partilha dos dados, quanto as práticas para a RDM, estão no escopo principal e influenciadas pelo acesso aberto sob a ciência aberta. Em sentido inverso, a partilha dos dados e as práticas para a RDM, requerem da ciência aberta, da abertura dos dados e do acesso aberto, a estruturação conceitual necessária para que o uso das práticas essenciais resulte em uma partilha adequada. A partir disso, assume-se que o nexos conceitual entre as práticas definidas como essenciais para a RDM e uma partilha adequada dos dados de investigação, envolve e passa inevitavelmente pela ciência aberta, sob o mote do acesso aberto.

Até este ponto, o construto apontou as relações conceituais em uma perspectiva existente entre dois eixos de estudo, as práticas para a RDM e a partilha dos dados de investigação, tendo a ciência aberta e o acesso aberto como nexos principal para as duas abordagens. Entretanto, e obviamente ainda sob uma perspectiva teórica, é necessário considerar os atores para as práticas e para que a gestão e a partilha dos dados sejam realizadas. Desta forma, o construto do nexos teórico torna-se completo quando, aliado ao vínculo conceitual entre os dois eixos de estudo citados, é inserido o fundamento teórico que permitiu identificar e articular os atores para a RDM na universidade, terceiro eixo de estudo.

O terceiro eixo de estudo tem por base as ideias dos autores discutidas ao longo dos capítulos 2, 3, 4, e principalmente o capítulo 5, que constituíram os atores e as articulações necessárias para a RDM no ambiente da universidade. Portanto, a seguir, são apresentados os resultados oriundos destes capítulos e o nexos teórico entre a articulação dos atores, a partilha dos dados e as práticas para a RDM. O nexos é apresentado de modo que a articulação dos atores no âmbito da universidade, possibilite a realização das práticas essenciais para a RDM e uma partilha adequada dos dados.

No capítulo 2, a relação entre a RDM e a universidade indicou que um caminho possível para identificar os atores desta relação é partir de classes genéricas, sugerida na tríade pessoas, políticas e infraestruturas. Estas classes juntas têm em sua base o desafio de gerir os dados de investigação e atender as demandas da ciência, sob a direção da ciência

aberta e do acesso aberto. No entanto, os resultados obtidos naquele capítulo apontam que, como tais atores estão inseridos em um ambiente complexo e heterogêneo, a identificação específica destes atores depende da exploração do contexto da universidade e das especificidades do ambiente da investigação. Nessa dualidade, as duas partes merecem igual atenção e os atores precisam ser identificados de acordo com o ambiente da investigação, onde as práticas essenciais para a RDM e uma partilha adequada dos dados possam ser realizadas de modo a considerar o contexto da universidade.

O capítulo 3 também apresenta uma série de atores humanos e não humanos inter-relacionados no ambiente para a partilha dos dados de investigação em acesso aberto. Tais atores são oriundos dos princípios e práticas da ciência aberta e foram agrupados em oito aspectos temáticos (Tabela 10) já aqui apresentados. Alguns dos atores que compõe estes aspectos ganham destaque quando examinados os inter-relacionamentos que os subsidiam. Assim, os atores humanos e não humanos destacados são os seguintes: a) os que compõem a comunicação científica; b) estabelecido pelas redes entre os investigadores e pelas redes entre os investigadores e as comunidades acadêmicas e entre ambos e a sociedade; c) os recursos humanos e materiais que possam apoiar a parte legal da investigação em termos de proteção, privacidade e segurança dos dados; d) a infraestrutura física e lógica, técnica e material para a abertura e sustentabilidade dos dados; e) as pessoas envolvidas no fluxo e na divisão do trabalho; e, f) a sociedade, através de cidadãos engajados com a ciência.

Ao observar os atores dos aspectos temáticos na relação com as práticas essenciais e na partilha adequada dos dados, verifica-se que não somente esta relação abrange os atores da tríade pessoas, políticas e infraestruturas indicadas no capítulo 2, mas a ultrapassa, tornando evidente as diferenças e a ênfase na necessidade da articulação. Assim, o resultado aponta para a exigência relacional entre os atores – aqui apresentados pelos aspectos temáticos dos princípios e práticas da ciência aberta e na partilha dos dados em acesso aberto. Esta exigência relacional (Figura 5) foi resumida nas seguintes afirmações: 1) uma partilha em acesso aberto não se completa de forma adequada sem um propósito, que seja lógico, organizado e ancorado nos princípios e práticas da ciência aberta; e 2) os aspectos temáticos constituídos pelos princípios e práticas da ciência aberta necessitam ser relacionados entre si para atingir a finalidade de uma partilha em acesso aberto. As afirmações também põe em evidência a

organização e o relacionamento fundados no acesso aberto e na ciência aberta, realizados por atores em uma dinâmica de articulações que convergem para a necessidade de sistematização.

No capítulo 4, o estudo dos modelos para a identificação das práticas essenciais para a RDM, a articulação e os atores também são evidenciados nos modelos estudados, pois estes apresentam em si as práticas para a RDM e para a curadoria, realizada por atores em uma dinâmica cíclica e processual. Os atores identificados no estudo que resultou nos elementos contextuais do modelo (Tabela 19) e nas práticas essenciais para a RDM (Figura 25), são os seguintes: a) o investigador como ator principal; b) os atores da comunicação científica; c) os recursos humanos; d) os materiais tecnológicos; e) os atores na parte da gerência do processo; f) os atores específicos para a operacionalização das práticas no domínio; g) os fluxos de trabalho, funções e responsabilidades; h) as comunidades envolvidas com a curadoria; i) a arquitetura, engenharia e as estruturas; e j) os atores das atividades de suporte e serviços. Alguns destes atores já foram aqui apresentados na concepção para a partilha dos dados, utilizando por vezes outros nomes e abordagens, isto indica que os atores oriundos dos modelos que resultaram nas práticas essenciais, também utilizam como seu fundamento os princípios e práticas da ciência aberta e o acesso aberto.

Em suma, os resultados dos capítulos de 2 a 4 indicam que os atores: 1) podem ser identificados através de classes genéricas; 2) precisam ser conhecidos de modo específico e de acordo com o ambiente da investigação; 3) existem em uma série de elementos humanos e não humanos, oriundos dos princípios e práticas da ciência aberta, que necessitam ser relacionados entre si para uma partilha adequada em acesso aberto; 4) estão articulados e constituídos sob práticas para a RDM, identificadas em uma dinâmica cíclica e/ou sistêmica, baseadas nos princípios e práticas da ciência aberta, no contexto do acesso aberto.

E por fim, a construção teórica no capítulo 5, através da teoria da atividade, fornece os principais resultados que permitem sistematizar os atores da RDM e possibilitar a adaptação ao ambiente da universidade. Os resultados oriundos deste estudo consideram que os atores aglutinados são estratégicos e operacionalmente interdependentes no processo articulatório, de modo a permitir através do sistema de atividade, a sua coordenação e arranjo por meio da sistematização. A partir disto, os atores envolvidos para a realização de uma partilha adequada dos dados de investigação, por intermédio de práticas essenciais para

a RDM, foram agrupados, articulados e sistematizados teoricamente (Figura 31) através dos seguintes atores: 1) sujeito (os investigadores); 2) as comunidades envolvidas e interessadas (agentes internos e externos à universidade); 3) as ferramentas necessárias (infraestruturas físicas e lógicas); 4) as regras (políticas locais, nacionais e internacionais); e 5) a divisão do trabalho (atribuições e funções). Todos estes atores, movidos por seus objetivos individuais, juntos e articulados estão direcionados ao cumprimento do objeto da atividade, qual seja: uma RDM realizada através de práticas essenciais para uma partilha adequada dos dados.

A articulação dos atores de modo a partilhar os dados e realizar as práticas para a RDM, está também relacionada à questão da complexidade e da heterogeneidade do ambiente. Esta questão, citada no primeiro capítulo, enfatizada no segundo e referenciada de modo recorrente nos capítulos posteriores, remete a necessidade de uma adaptação dos atores articulados no ambiente e ao contexto da RDM na universidade. É sob essa necessidade que os resultados para a articulação dos atores apresentou uma proposta de ciclo metodológico (Figura 31) que permite uma abordagem resultante na sistematização e adaptação contextual, através de um método de intervenção formativa de aprendizagem, constituída por seis fases, quais sejam: 1) sistematização; 2) enquadramento; 3) avaliação; 4) concepção; 5) ensaio; e 6) efetivação. Tal adaptação possibilita que atores articulados e sistematizados, realizem as práticas essenciais para a RDM que resulte em uma partilha adequada dos dados, de acordo com o ambiente e no contexto da universidade.

Portanto, com base nos resultados apresentados, pode-se definir a articulação dos atores neste estudo como: um tipo de articulação sistêmica que possibilita a realização de práticas essenciais para a RDM e uma partilha adequada dos dados, através do relacionamento entre os elementos humanos e não humanos, constituídos a partir dos componentes envolvidos para a RDM na universidade, sob o acesso aberto e fundamentado na ciência aberta. Para tanto, os atores são agrupados por classes de investigadores, comunidades, ferramentas, políticas e atribuições para o trabalho, viabilizados por um método interventivo de aprendizagem que permite a adaptação ao ambiente, mediante um processo onde os atores são sistematizados, os problemas enquadrados e avaliados para conceber soluções, que após modeladas e postas em testes podem ser efetivadas para o uso, ou retornar ao processo para aperfeiçoamento. Quando efetivado, o processo resulta em

atores articulados e uma RDM adaptada ao ambiente, para a realização de práticas essenciais de gestão e uma partilha adequada dos dados de investigação.

Os atores da RDM na universidade e suas possíveis articulações, envolvem várias atividades que abordam e são influenciados por um grande conjunto de fatores de difícil realização, pautados pelas características do próprio ambiente (Pinfield, Cox & Smith, 2014). Neste sentido, as práticas realizadas em tal contexto de atores em uma infraestrutura convencional também acompanham as características ambientais (Slota & Bowker, 2017). No entanto a necessidade de articulação dos atores (Whyte & Tedds, 2011), faz com que uma RDM que pretenda realizar práticas essenciais para resultar em uma partilha adequada dos dados, deve estar sob um forte vínculo conceitual, para que também possa direcionar e suportar as complexidades e heterogeneidades do ambiente da universidade.

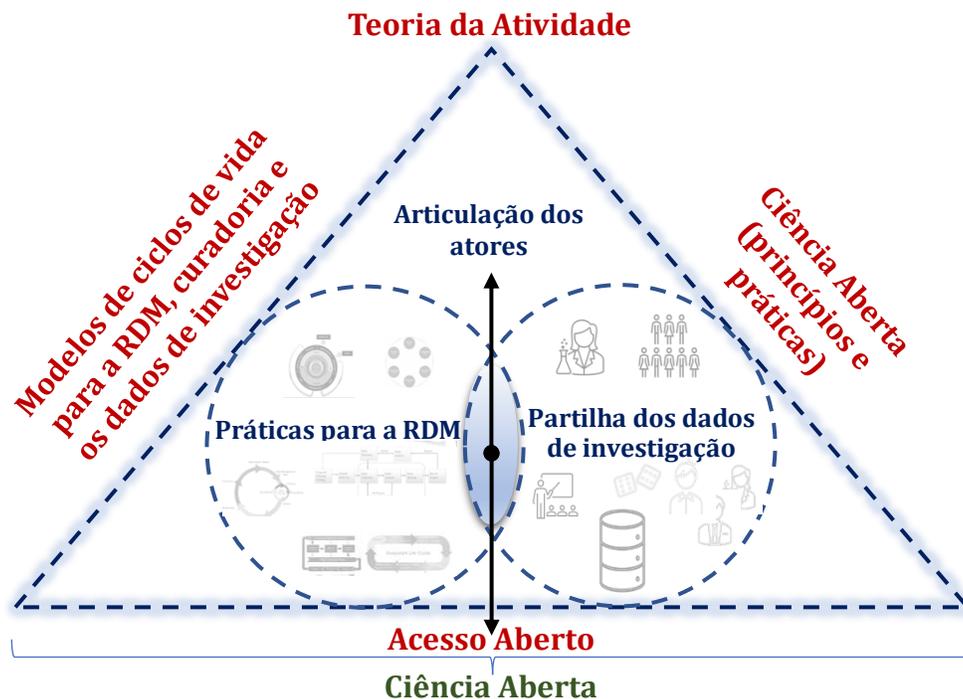
Sob a necessidade referida anteriormente, verifica-se que a base conceitual resultante para a articulação dos atores, também tem na ciência aberta e no acesso aberto o nexos teórico para o relacionamento com os outros dois eixos de estudo – a partilha dos dados de investigação e as práticas para a RDM. Além disso, percebe-se ainda que estes dois eixos estão unidos igualmente pelas articulações dos atores, visto que tais atores são comuns e necessários para que as práticas para a RDM e a partilha dos dados sejam realizadas. Assim, aponta-se que a articulação dos atores constitui-se também como elemento integrador na realização das práticas para a RDM e para a partilha dos dados.

No entanto, identificar o nexos não parece ser tarefa muito difícil quando os resultados provenientes da literatura são aqui apresentados, talvez a dificuldade maior esteja na organização para que o vínculo dos três eixos conceituais, sejam utilizados de modo prático para exercer as suas funções em conjunto e com um propósito único e direcionado. Portanto, além do nexos teórico a seguir representado (Figura 32), os resultados reunidos são utilizados para a fundamentar o modelo, bem como indicar uma utilização prática para a RDM no âmbito da universidade, com atores articulados, que possibilite a realização de práticas essenciais para a gestão e a partilha adequada dos dados. Nesse sentido, o nexos aqui apresentado sustenta conceitualmente o modelo proposto e aponta o caminho para a sua prática.

É possível concluir que a análise do resultado da literatura aqui realizada, procedente dos três eixos de estudo - a partilha dos dados de investigação, as práticas para a RDM e a articulação dos atores -, fundamentada respectivamente na ciência aberta, nos modelos de ciclos de vida e na teoria da atividade, resultou nas definições apresentadas sobre a partilha adequada dos dados de investigação; as práticas essenciais para a RDM e a articulação dos atores, todas observadas para o ambiente da universidade. Os resultados são importantes tanto para o confronto com os dados empíricos, quanto para a fundamentação do modelo proposto. Percebe-se ainda, que a relação de interdependência e complementaridade entre os eixos de estudo, tem seu nexos e interseção no acesso aberto fundado na ciência aberta. Infere-se também que, na perspectiva do estabelecimento desta relação conceitual, os dois primeiros eixos estão de forma inexorável e funcionalmente unidos também sob a articulação dos atores da RDM na universidade. Assim, pelo fato destas relações não estarem explicitamente construídas na literatura, é oportuno sugerir uma proposta para a relação conceitual entre as práticas para a RDM, a partilha dos dados de investigação e a articulação dos seus atores, conforme constantes na figura abaixo.

Figura 32

Relação conceitual entre as práticas para a RDM, a partilha dos dados e a articulação dos atores



É importante observar que a relação proposta também resulta no modelo conceitual desta investigação. Assim, é necessário que sob o nexu teórico apresentado, a articulação dos atores, as práticas essenciais para a RDM e a partilha adequada dos dados de investigação, oriundas do estudo dos eixos teóricos e representados através das definições aqui elaboradas (Tabela 23), sejam associadas à análise e a discussão dos dados qualitativos provenientes do campo empírico, apresentadas na próxima seção. Esta associação permite que os dados coletados por meios das entrevistas, sejam confrontados e discutidos com o construto da análise teórica realizada, de modo a apontar caminhos para conceber o modelo proposto para a RDM no ambiente da universidade, objetivo desta investigação.

6.2. AMBIENTE E CARACTERÍSTICAS DOS DADOS E DA INVESTIGAÇÃO

O primeiro bloco de perguntas refere-se à identificação prévia da investigação e dos atributos dos dados da investigação, tendo a intenção de perceber o contexto em que os dados são administrados, bem como obter informações relacionadas aos aspectos e utilidade destes objetos, a partir das Unidades de Investigação e Desenvolvimento (UI&D) da Universidade de Coimbra (UC). Retrata-se com isso tanto o contexto, quanto as características dos dados geridos pelos investigadores no ambiente da investigação.

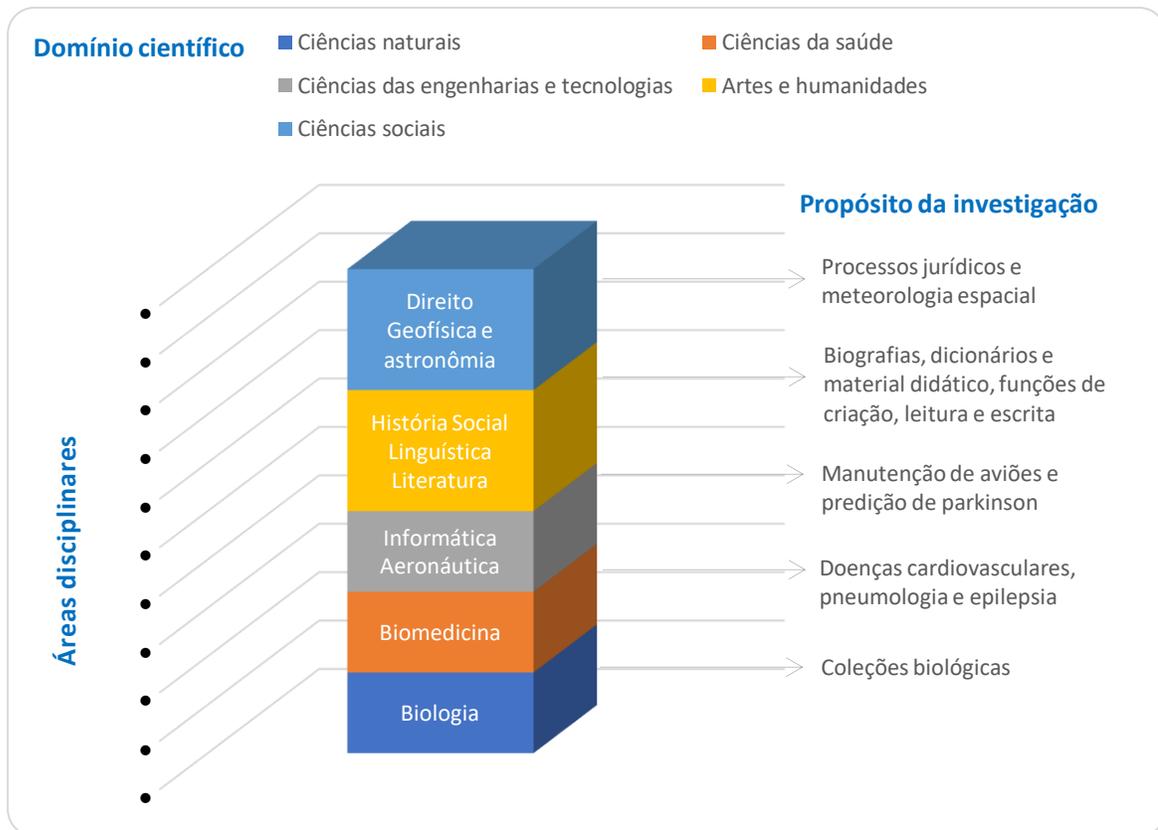
6.2.1. AMBIENTE E PROPÓSITO DA INVESTIGAÇÃO

Ao serem inquiridos sobre o propósito da investigação, o domínio científico e as áreas disciplinares na qual os dados foram ou estão sendo geridos, as respostas dos investigadores, com dados administrados em áreas disciplinares de cinco dos seis domínios determinados (Tabela 5)⁴⁸, são caracterizadas pela diversidade do ambiente e de propósitos das investigações, conforme apresentado na figura 33 e nas informações que a ela segue.

⁴⁸ Conforme exposto na abordagem metodológica, constante no primeiro capítulo, as áreas disciplinares e os domínios científicos foram determinados de acordo com o que é adotado pelo Instituto de Investigação Interdisciplinar (III) da Universidade de Coimbra (UC), tal Instituto também viabilizou o início deste trabalho empírico, fornecendo uma lista com o nome das Unidades de Investigação e Desenvolvimento (UI&D) sob seu escopo de atuação, com as possíveis características a serem exploradas em relação à RDM, juntamente com o contato dos responsáveis pelas UI&D listadas. Assim, tendo em vista o interesse na observação da diversidade, foram convidados para as entrevistas, que resultaram nos dados aqui apresentados, investigadores pertencentes aos seis domínios científicos definidos de acordo com as categorias do III da UC.

Figura 33

Visão do domínio científico e das áreas disciplinares em relação com o propósito da investigação.



Nas explicações dos respondentes, a diversidade ambiental é percebida pelo amplo espectro de propósitos e formas de investigações identificadas, que variam desde a investigação sobre “plantas (...), fungos e líquens” (E1), passando por “questões da segurança social” (E2), até o uso de “aplicações (...) em *smartphones* para pessoas com Parkinson” (E5). Conforme constatadas também nas transcrições abaixo:

Uma coleção biológica (...) é uma investigação sobre plantas (...), também se encontram alojados organismos que sofrem um tratamento semelhante (...), são fungos e líquens (E1).

São projetos de investigação em equipa, mas também projetos de investigação pós-doutorados, projetos de investigação doutorado (...). A maior parte da investigação (...) é (...) em ciências sociais aplicadas (E3).

Na astronomia, principalmente no sol (...) e dados geomagnéticos (...), os investigadores do centro trabalham muito esse tipo de dados (...), mas não só (E4).

(...) aplicações que corresse em *smartphones* para pessoas com Parkinson (...). O que estávamos a tentar perceber era quantos toques podia uma pessoa com Parkinson fazer em conforto, sem entrar nas questões do *freezing* (...). Criamos a sensação visual para as pessoas (E5).

Uma empresa da aeronáutica queria desenvolver (...) algoritmos para manutenção preditiva dos aviões (...). Ao invés de ter uma data fixa (...), os algoritmos (...) permitem prever a altura ideal para fazer manutenção (E6).

História social (...). O objetivo é a criação das bases de dados com as quais as comunidades possam interagir (...), consultar, fazer sua genealogia (...) voltada para reconstituição de família, histórias de vida, biografias (...), percursos de vida, perceber as suas escolhas (...), um mapeamento da estrutura social local, como ela foi evoluindo num determinado ponto de tempo (E7).

É um projeto internacional (...) de *crowdsourcing* (...). Nós estamos criando um jogo para coletar (...) bons exemplos ou conteúdo sensível (...) em corpus *Web* para fins didáticos (...), seja para uso em dicionários (...) ou em material didático. Na nossa equipe nós somos representantes de cinco línguas diferentes: português, esloveno, estoniano, holandês e sérvio (E8).

Eu tenho trabalhado principalmente com dados clínicos (...), cardiovascular (...), pneumologia (...). Em ambiente hospitalar ou medidos em casa (...). Envolve a predição de crises epiléticas (E9).

Este projeto começou por ser um arquivo e transformou-se (...) [em] um espaço criativo de exploração, de especulação, de investigação (...) para simular certas funcionalidades (...) como as funções de ler, escrever e editar (...) e a possibilidade de reescrita do texto [literário] (E10).

Serve para resolver um problema que é, quando se estuda pessoas na história, as informações sobre as pessoas estão dispersas em muitas fontes, em muitos documentos diferentes (...), a informação que temos sobre as pessoas varia com o tempo (...), neste tipo de investigação é importante perceber a necessidade de preservar os dados de uma forma muito próxima ao original (...). [O sistema informático] foi desenhado com essa filosofia já a muitos anos atrás (E11).

Este grupo de investigações, caracterizado por domínios distintos e áreas disciplinares diferentes, demonstra a pluralidade de temas e práticas com que os dados de investigação precisam ser geridos. Trata-se de formas, objetivos investigativos e abordagens muito diversificadas entre si, como a manutenção de aviões, biografias, dados clínicos de doentes, corpus *Web* ou investigações sobre plantas, sol, ou simulação de funcionalidades literárias, mas também processos que envolvem algoritmos, *smartphones*, mapeamentos, sistemas informáticos e *crowdsourcing*. Embora pequena, a amostra representa a diversidade possível que acontece no ambiente da RDM na universidade.

A constatação de que os dados gerados nas investigações são destinados a diferentes propósitos, em diferentes disciplinas e áreas científicas, por meio de abordagens distintas, intensifica a percepção da diversidade no ambiente da gestão dos dados (Sayão & Sales, 2019). Borgman (2012) explica que essa diversidade no contexto da investigação, quando observados os diferentes métodos e abordagens em que os dados estão envolvidos, facilita a entender as condições sob as quais os investigadores estão dispostos a partilhar “seus dados”. Tal afirmação é um aspecto importante a ser verificado no ambiente, porque

direciona o olhar para o investigador, com a finalidade de compreender o seu ambiente científico e disciplinar, juntamente com as influências exercidas sobre ele para a partilha.

Este contexto diverso também pode sugerir a utilização de uma multiplicidade de modelos e de abordagens específicas para que a RDM possa ser mais efetiva em atender às necessidades dos usuários (OCDE, 2007). No entanto, alerta-se que abordagens muito específicas são propícias a resultar em excessos de especialização e compartimentalização, que pode anular o encorajamento da investigação interdisciplinar e a interpretação dos dados em diversos contextos, um dos benefícios esperados para a RDM (Sayão & Sales, 2019).

A compreensão das especificidades das investigações no ambiente distribuído por áreas e disciplinas distintas, é fundamental para poder gerir os dados, articular os atores e utilizar práticas essenciais para a gestão e a partilha adequada dos dados. É necessário ter em conta que na definição de uma partilha adequada, o investigador é o ator principal, e, como tal, deve ser observado sob suas influências contextuais que repercutem nas práticas e na gestão para a partilha dos dados. Logo, a influência exercida pelo ambiente disciplinar da investigação sobre o investigador, é parte importante da análise a ser considerada na RDM para a partilha. É interessante também observar que, ao mesmo tempo em que há a necessidade de atentar para as especificidades do ambiente, não é viável criar modelos para cada realidade. Assim, o modelo aqui resultante deve suprir tal dificuldade, tendo um viés generalista, sem deixar de ser flexível para considerar as especificidades e as diferenças disciplinares (Lämmerhirt, 2016; Severin et al., 2018). A importância destas soluções são asseveradas quando se percebe que este ambiente diverso, possivelmente seja uma fonte geradora das práticas isoladas para a RDM.

Ainda sobre o ambiente, os entrevistados puderam responder acerca da capacidade da condução da investigação, realizada especificamente no contexto das UI&D. Os investigadores destacaram as dificuldades de recursos, a sobreposição de influências em suas atividades e as diferenças de capacidades que envolvem a RDM no âmbito em questão. As respostas seguem de acordo com as citações abaixo:

Trabalhavam aqui dez investigadores (...), o número de pessoal, de investigadores a pessoal de manutenção, foi reduzido em 83,4% (...), ou seja, fiquei eu (...) e uma funcionária, estivemos assim durante onze anos (...), entretanto houve uns projetos (...) a nível nacional (...) que alimentaram de

bolsheiros durante algum tempo. (...). Queriam que fossemos uma estrutura museológica (...), é uma situação atípica (...) que deixa (...) altamente vulnerável (...). [A unidade] é mais um recurso da universidade do que uma estrutura transitória (E1).

[A UI&D] foi evoluindo em termos de aumento da sua capacidade, não é ainda uma capacidade extraordinária, mas se compararmos a muitos centros de investigação está na primeira linha das práticas de investigação, pela dimensão, pelas capacidades financeira, científicas, infraestruturais etc. (...). Em Portugal temos alguns centros de investigação de grande dimensão na área das ciências sociais e humanas, mas são poucos, e depois temos muitos pequenos centros de investigação que não têm essa capacidade e, portanto, essa é uma realidade que está longe de ser efetiva. Diria que temos investigação a várias velocidades, por diferentes capacidades financeiras, infraestruturais, de recursos humanos qualificados etc (...). Não é fácil (...) que as instituições tenham recursos humanos e financeiros que possam investir (...) para garantir que haja mais informações canalizadas para estas infraestruturas da partilha da informação (E2).

(...) temos uma estrutura (...), uma equipa de recursos humanos (...) que se dedica completamente à gestão de projetos e que ganhou já uma experiência ao longo dos anos em todos esses assuntos (...), de facto, instituições mais pequenas com estruturas mais pequenas não têm esta capacidade (E3).

(...) [é] um centro novo (...), tem vindo a latejar em termos de recursos humanos, nós tínhamos uma [pessoa da área administrativa] para os dados solares, entretanto (...) reformou-se e não pôs-se mas nenhum (...). Não há um quadro funcional de trabalhadores (...), há uma secretária e um técnico superior, são as únicas duas pessoas (...). Antes tinha herdado quatro, entretanto uma reformou-se, outra reformou-se (...) agora só tem dois (...). Há centros que tem um peso relativo muito maior na universidade (...) pelas áreas que trabalham e pelos pesos que aglomeram e pelos dinheiros que gerem e temos que olhar para todos com os mesmos olhos (...), mas com as diferenças que eles têm (...). As universidades (...) em geral (...) olham pela ótica empresarial dos números, sucesso (...), regem-se por métricas muito empresariais (E4).

A diversidade já observada no ambiente das investigações, também aparece nas UI&D, de modo que os relatos indicam dificuldades e diferenças entre unidades de investigação. Essas dissemelhanças organizacionais, segundo os entrevistados, a depender da capacidade da UI&D em que estejam alocados geram dificuldades estruturantes. São indicados problemas de variação ou falta de recursos humanos, financeiros e infraestruturais, mudanças na organização e tratamentos diferenciados provenientes de instâncias superiores, que afetam o ambiente da investigação, de forma a gerar vulnerabilidades e influências na capacidade evolutiva das unidades de investigação.

Os apontamentos extraídos das entrevistas estão em consonância com relatório da Comissão de Investigação do Conselho Geral da Universidade de Coimbra, de novembro de 2019. No relatório é expresso que 81% das UI&D citaram os recursos humanos como maior fator de constrangimento. Apresenta também a existência de unidades de dimensão relativamente grande, que concentram muitos dos investigadores de uma área, enquanto, no extremo oposto, alguns centros representam a menor proporção. O documento também conclui que “a disparidade de dimensão encontrada é uma situação que mereceria ser

debatida”. O relatório aponta claras desigualdades entre as UI&D integradas na UC e as UI&D autônomas, gerando “sentimentos mútuos de falta de equidade que urge minorar”. Em termos interdisciplinares, o documento indica que na UC, entre as diversas UI&D, observa-se uma articulação deficitária sendo importante reunir “boas práticas”, e ajudar a torná-las mais transversais a todo o universo da instituição (Relatório UC, 2019, pp. 9; 50; 52).

As situações e capacidades provenientes do ambiente das UI&D em que a investigação está alocada, remetem às dificuldades e diferenças entre as unidades. Tais elementos constituem outro fator influenciador importante a ser observado para que a articulação dos atores no âmbito da universidade, possibilite a realização das práticas essenciais para a RDM e resulte em uma partilha adequada dos dados. Aspectos como a infraestrutura física, os recursos humanos e financeiros e os sentimentos de insatisfação ou satisfação, são gerados a partir de tal ambiente e afetam a RDM aqui objetivada. No entanto, percebe-se que, além das dificuldades apresentadas, existe também o relato de condições adequadas para condução da investigação devido à capacidade da UI&D, o que do mesmo modo caracteriza as diferenças e podem reforçar os problemas apresentados.

Tomada por referência a afirmação da Fundação para a Ciência e Tecnologia (FCT, 2021), de que é possível por intermédio das UI&D promover o desenvolvimento de melhorias nas dimensões institucional, econômica, sociocultural, territorial e ambiental, então infere-se que a minimização das dificuldades e das diferenças relatadas, são medidas necessárias para a realização da RDM aqui objetivada. Sem a efetivação dessas medidas, as diferenças verificadas entre as UI&D podem diminuir a possibilidade da interdisciplinaridade (OCDE, 2021; Relatório UC, 2019; Sayão & Sales, 2016) nas investigações a partir da RDM, e ainda, reforçar um afastamento gerado pelas disparidades de infraestruturas, recursos disponíveis e sentimentos gerados que tendem a repercutir para o isolamento dos dados. São aspectos sociotécnicos que devem ser observados, com situações geradoras de tensões e conflitos, que impedem o avanço do tipo da RDM aqui proposta.

Se as dificuldades relatadas forem comparadas com o Relatório da UC (2019), é expectável que seja desenvolvida um tipo de articulação entre as UI&D que compreenda atores e práticas, onde os recursos informacionais possam ser partilhados para prover serviços que minimizem as diferenças na capacidade operacional em relação a RDM. Tal

solução vai ao encontro das recomendações do relatório, que apontam os seguintes direcionamentos: a) a articulação entre as unidades de investigação carece de sistemas de informação que respondam à necessidade de informação de docentes e investigadores, com potenciais colaborações e valências; b) recomenda-se a articulação dos seus centros de investigação, promovendo a melhoria da qualidade das estruturas de investigação e a motivação generalizada; e, c) se elege a interdisciplinaridade como uma vertente de trabalho importante. Em outro sentido, quando os dados aqui coletados são comparados com o resultado da análise da literatura, ratifica-se a ideia de que as práticas essenciais para a RDM devem ser associadas ao ambiente e ao conhecimento contextual, pondo-se a questão da influência ambiental em foco, desta vez direcionado ao ambiente das UI&D, mas que, de forma importante, pode interferir na RDM para uma partilha adequada dos dados.

As duas questões desta seção resultam na observação do ambiente pela lente da diversidade, primeiro na investigação, através das especificidades disciplinares e domínios diversificados e, depois nas UI&D, pelas diferenças e problemas estruturais apresentados. Os dois tipos de diversidade e as suas consequências são postas como determinantes em relação ao modelo para a RDM, porque afetam a capacidade da investigação e as práticas do investigador, bem como a articulação desejada para a universidade. Essas dificuldades são um campo propício para que os dados se tornem invisíveis através de práticas isoladas, pois as peculiaridades de cada investigação e as dificuldades do ambiente podem fazer com que o investigador pense, aja ou partilhe de forma individualizada e colabore para maximizar os problemas relatados.

Portanto, a articulação dos atores e o uso de práticas essenciais para que a RDM partilhe adequadamente os dados neste contexto, requer primeiramente a compreensão do ambiente disciplinar em seus diversos tipos de investigação, mas também do ambiente institucional, nas necessidades apresentadas pelas UI&D, sem perder o olhar para as influências que ambos possam causar no investigador. Neste sentido, o modelo proposto tem dois aspectos a serem considerados, a preocupação com o individual e com o institucional. Ou seja, deve-se abordar o ambiente dos dados de investigação em uma concepção multidimensional (Borgman, 2015).

6.2.2. CARACTERÍSTICAS DOS DADOS DA INVESTIGAÇÃO

Nas perguntas sobre o tipo, o formato e o volume dos dados geridos, pôde-se através das respostas dos investigadores apresentar a caracterização dos dados, remetendo aos atributos que indicam a fonte, a natureza e a sua materialidade, conforme exibido na figura abaixo. As respostas em seu contexto estão dispostas nas transcrições que seguem a figura.

Figura 34

Caracterização da origem, fonte e materialidade dos dados de investigação



A coleção tem para cima de oitocentos mil itens [físicos] (...), mais de noventa mil registos (...) em formato digital (...), são dados textuais (...) e imagem (...). O material (...) se consegue transformar numa versão digital (E1).

São dados observacionais (...), hoje em dia tem muita coisa digitalizada (...), já não é em chapa fotográfica (...) as imagens do sol e do espectro solar (...), mas ainda fazemos observações em suporte físicos (...) e temos para trás uma série de dados de registo em papel (E4).

(...) os dados [que] eram recolhidos (...) corriam no ecrã do telemóvel (...), tempos, locais de toque, precisão (...), fizemos (...) interfaces (...) divididos em secções (...) de diferentes tamanhos (...), com um alvo que era preciso levar do lado esquerdo para o lado direito (E5).

Tipicamente são dados técnicos (...), os aviões têm muitos sensores internamente (...), durante o voo esses aviões estão constantemente a recolher informações e a gravá-las no sistema de bordo

(...). No avião esses dados são coletados automaticamente (...), depois de devidamente analisados com algoritmos de inteligência artificial (...) permitem prever (...) [a] manutenção (E6).

Estamos a falar de quatrocentas mil ocorrências (...) [em umas das] base de dados (...). Essencialmente [são] informações sobre pessoas (...). Nos arquivos hoje em dia, muitos (...) já estão sob a forma de objetos digitais (E7).

Esse jogo ajuda a buscar bons exemplos (...) em corpus *Web* (...). O que nós queremos é que os jogadores nos ajudem a identificar bons exemplos ou conteúdo sensível ou vulgar ou com algum tipo de problema na estrutura que eles achem que não seja adequado para fins didáticos (E8).

São estudos observacionais (...). Nós temos trabalhados mais com sinais (...), eletrocardiograma, sons cardíacos, sons do pulmão (...), eletroencefalograma (...), as vezes (...) usamos imagens. Se estivermos a falar (...) [da investigação sobre] epilepsia, são quantidades enormes de dados (...), são aquisições de dados durante (...) meses (...). Nos projetos típicos (...) [há] bases de dados com mil, dois mil (...), dez mil doentes. Trabalhamos uma vez com uma base de sessenta mil (...), já são valores enormes (E9).

Podemos falar aqui (...) de dois tipos de dados diferentes (...), aquilo que é (...) associado a conta gmail (...), são dados estatísticos (...). Depois temos um outro nível, que são os dados que os utilizadores nos dão (...) para a utilização no âmbito das funcionalidades do site (...). Não é assim muito grande, a maioria dos ficheiros são textos (...), não ocupam assim muito espaço (E10).

Quando recolhemos a informação, não sabemos quem é quem (...), só sei isso quando [se] recolhe muita informação (...) [para] depois distinguir os dados em sua forma original, daquilo que vão ser posteriores decisões sobre os dados (...), separar o momento da recolha ou da observação do momento da interpretação. Pôr [os dados] no *GitHub* (...) dizer as pessoas para ir lá (...) e descarregar meus dados, todos primários que eu recolhi na minha investigação (E11).

É possível perceber diretamente nas transcrições ou interpretar pelo contexto, que os dados geridos nestas investigações foram relatados em formatos variados: são imagens, papéis, objetos de coleção, sons, sinais ou textos, caracterizados como materiais físicos ou digitais. Na investigação pode-se ter que gerir ambos os materiais, analógicos e digitais, como relata o E4: “hoje em dia tem muita coisa digitalizada (...), mas ainda fazemos observações em suporte físicos”. Observa-se também que há um processo de transformação do físico em digital: “O material (...) se consegue transformar numa versão digital” (E1). “Nos arquivos hoje em dia, muitos (...) estão sob a forma de objetos digitais” (E7). O volume de dados digitais mencionados também varia de acordo com o tipo de investigação e os seus quantitativos, vão desde bases de dados com sessenta, noventa mil registros, até textos que não ocupam muito espaço. Para o processo da coleta também se verifica que são utilizados diferentes instrumentos, como os que procedem de aviões, de toques em telemóveis, de aparelhos sensores utilizados no corpo humano, arquivos públicos ou funcionalidades dos sistemas alocados na Internet. Tais dados têm a sua natureza originária possibilitada pela tecnologia computacional, mas também de modo observacional e registos de caráter variados.

É esperado que o contexto diverso da investigação na universidade, naturalmente produza dados com características variadas, o que assevera a possibilidade de fragmentação, invisibilidade e isolamento que caracteriza este ambiente. No entanto, embora relacionada com os problemas referidos, a diversidade é observada aqui por outro viés – o da heterogeneidade dos dados de investigação. Os dados, como objetos informacionais centrais para a RDM, devem ser assim observados, pois desta forma pode-se perceber com mais exatidão o valor que eles possuem, como e onde armazená-los, possibilitando saber o que fazer com eles e mesmo os danos que as suas perdas podem causar, como no caso dos dados qualificados como observacionais, caracterizados como irrecuperáveis (NSF, 2005).

A heterogeneidade dos dados no tocante a formatos, mídias, suportes, expressões, arcabouços tecnológicos, estruturas e complexidades, é caracterizada como parte das dimensões da diversidade pertencentes a universidade (e-IRG, 2016). Nesta dimensão, os dados, com volumes, formatos e natureza distintos, estão em todos os lugares no ambiente e não são fáceis para gerenciar e partilhar. Como indicam Hedstrom e Myers (2014), a RDM e a partilha têm emergido como um grande desafio, pois os dados estão em formatos materiais de difícil integração e são obtidos em processos variados, mas que, se explorados apropriadamente, podem gerar benefícios. Logo, é necessário que os dados gerados neste ambiente, sejam caracterizados, estudados e organizados para evidenciar os benefícios da ciência aberta para a RDM (Masuzzo & Martens, 2017).

Gray et al. (2002), afirmaram que os dados se tornam incompreensíveis e, portanto, inúteis, a menos que se tenha uma descrição detalhada e clara de como e quando foram obtidos e de como foram produzidos, ou seja, é necessário além de caracterizá-los, os contextualizar. Borgman (2010), diz que a caracterização dos dados de investigação é passível de interpretações e reinterpretções. Portanto, juntamente com a caracterização, os dados devem ser acompanhados de uma leitura do ambiente e dos seus significados.

No confronto com os resultados teóricos, confirma-se que, para que os dados sejam adequadamente partilhados, deve-se prover documentações que facilitem o seu uso atual e futuro. Assim, é preciso realizar práticas de modo a conhecer os processos, formatos, volumes, fontes e outras características dos dados que podem ser diferentes entre si, mas

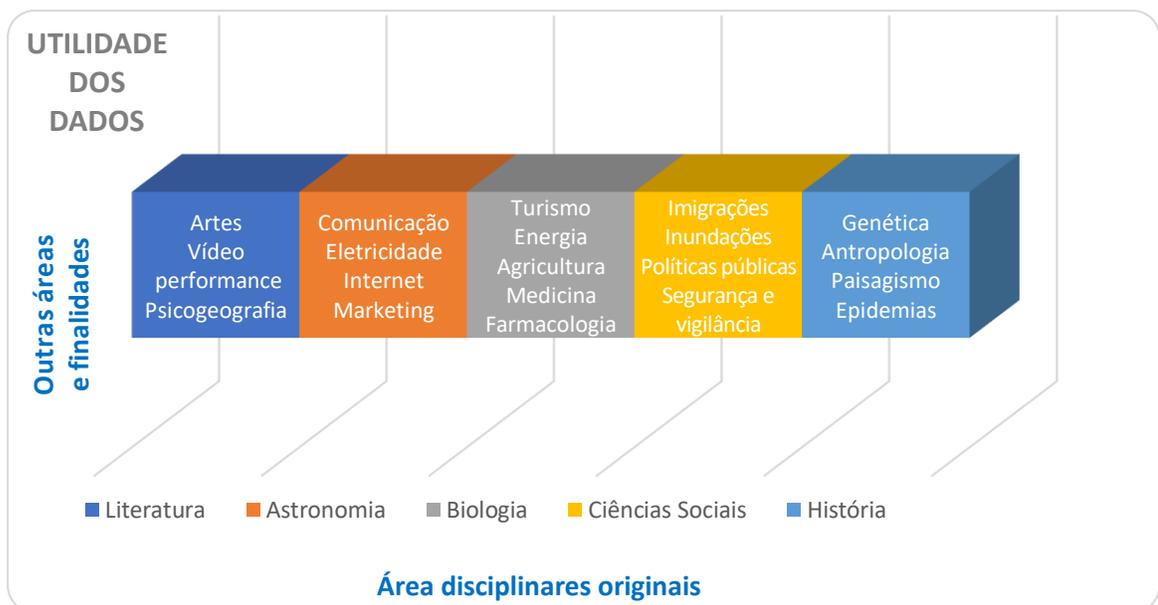
necessários para garantir a sua descrição e a do contexto, bem como o dimensionamento dos seus atores resultem em articulações e práticas para a gestão e a partilha adequada dos dados.

Portanto, a heterogeneidade dos dados irá demandar estratégias e suportes gerenciais, infraestruturas tecnológicas, políticas, múltiplas interações, cuidados e investimentos em escalas bastantes distintas para sustentar as diferenças, os processos e a natureza diversa dos dados. Consequentemente é possível pensar que articular atores com práticas necessárias para gerir volumes, tipos e formatos de dados tão distintos, no caminho para uma partilha adequada, torna-se uma tarefa inalcançável. No entanto, para que a RDM aconteça de modo relevante nesse contexto, considerando a complexidade e as dificuldades do ambiente, bem como a necessidade de recursos coordenados, um caminho possível é o seu desenvolvimento gradual, ou seja, alcançando primeiramente os itens comuns e mais simples, seguindo sucessivamente para uma maior complexidade e abrangência. Assim, a partir da sua caracterização, os dados podem ser organizados através de práticas essenciais para a RDM e por atores articulados, tendo como resultado uma partilha adequada.

Perguntados sobre a utilidade dos dados para uso em outras áreas e finalidades distintas das quais foram originalmente planeados, os respondentes indicaram o uso e possíveis usos de modo interdisciplinar e social, conforme a figura 35 e as citações seguintes.

Figura 35

Utilidade dos dados em áreas e finalidades diferentes das quais foram planeados.



Na realidade esses dados são relevantes para uma enormidade de áreas científicas (...), medicina, farmacologia, agricultura, energia (...) e como fonte de medicamentos (...), é aí que vão buscar a base do conhecimento (...) e depois vão ser utilizados (...) até pelo turismo para escolherem as zonas para fazerem passadiço, para saberem os sítios melhores para levarem as pessoas (E1).

Nas alterações climáticas (...), [deve-se] tratar dos riscos ambientais, mas também dos riscos sociais (...), se há uma questão sobre inundações, as inundações vão afetar as populações (...). As práticas para garantir que o risco [seja] minimizado numa população costeira, implica o envolvimento das ciências sociais (...). Nas questões das imigrações, do controlo de fronteira dos fluxos migratórios, onde há questões de segurança e vigilância (...) não se resolve sem (...) participação das ciências sociais e humanas (E2).

Nós trabalhamos numa área fundamental e que agora nos tempos da pandemia se tornou importantíssima (...), [por exemplo] estamos a falar agora (...) via eletricidade e via Internet e se houver um problema grande em termos físicos (...), a nossa monitorização (...) [ajuda a] prevermos, eventualmente, alguns fenómenos grandes que podem pôr em causa precisamente a rede elétrica, a rede *Wi-Fi*, a rede de Internet e a rede de satélites (...), um grande evento solar, uma tempestade solar que interfira com a nossa rede elétrica (E4).

Imagino que possa ser usado para mais coisas (...), são dados de operação de uma companhia aérea (...), portanto imagino que haja muitas análises a serem feitas com base naqueles dados (E6).

Este tipo de dados tem características muito específicas, a única outra área que eu já vi dados semelhantes serem tratados é na genética (...), mas eventualmente informações que constem deles podem ter interesse para outras áreas (...) [como] situações epidêmicas (...) doutros tempos que podem ajudar a perceber a causa de morte (...), às vezes a antropologia pode ter interesse nestes tipos de dados (...). Os dados históricos (...) estão abertos a questionamentos de outras áreas que queiram dar uma espessura temporal as suas investigações (...), até em termos de paisagem do território, das organizações dos poderes (...), nada impede que a geografia, a antropologia e a genética possam olhar para eles, porque as origens interessam sempre para compreensão de um aspecto presente (E7).

Nós temos várias pessoas a trabalhar na Colômbia, Estados Unidos em Espanha em Portugal (...) no Brasil, em contexto de sala de aula em várias universidades (...), temos a utilização do arquivo como espaço de leitura (...). Em contexto de trabalho artístico (...), houve um grupo de estudantes da universidade de (...) Salamanca que fez um conjunto de vídeo-performances (...), fizeram uma edição virtual (...) que funcionou como um roteiro (...). Podemos imaginar [o uso] (...) a nível de psicogeografias (...), é possível imaginar (...) o mapa de Lisboa com as referências topográficas que estão na obra (E10).

Há um caso de aplicação, absolutamente espetacular (...) que acho que teria um impacto muito grande que é o ficheiro dos alunos da Universidade de Coimbra, desde 1537 (...) até 1913, que é quando a universidade de Coimbra deixa de ser a única universidade de Portugal e também do espaço de língua portuguesa (...), tá ali toda a gente que teve uma educação superior no espaço de língua portuguesa (...), todos os administradores coloniais (...), todos os que vinham das colônias (...), é um interesse global (...), é um tipo de dados que toda gente quer integrar em qualquer tipo de investigação (...) seja o que for (E11).

As respostas fornecidas podem ser analisadas em duas perspectivas. A primeira toma por base as transcrições onde os investigadores efetivamente utilizam os dados ou citam a possibilidade do uso em campos distintos, como as que ligam a astronomia ao *marketing* institucional, a literatura à vídeo performance, a biologia ao turismo, as ciências sociais à vigilância e a história à genética. Embora um dos respondentes pareça ter sido

genérico: “Imagino que possa ser usado para mais coisas” (E6), na maioria dos relatos, a percepção de utilidade aponta para mais de uma finalidade em áreas aparentemente distintas, sendo tais utilizações mais ou menos evidentes, a depender da perspectiva do observador que a define. Assim, um fator importante para partilhar adequadamente os dados, está em definir onde eles podem ser utilizados, visto que a definição resultante da análise da literatura para este tipo de partilha, aponta que todo o processo deve ser objetivado para a reutilização dos dados. A possibilidade do uso ou reuso em outros campos é necessário e importante como um dos objetivos da RDM, podendo resultar no alcance da interdisciplinaridade e na utilização pela sociedade com vistas ao benefício social.

No entanto, segundo Borgman (2007), é desafiador tornar o conteúdo criado para uns, útil para outros. Nesse sentido, uma prática possível na superação do desafio de tornar os dados úteis, é o próprio investigador atribuir onde tais dados poderiam ser utilizados, pois quem os conhece originalmente tem a percepção da sua utilidade. Por vezes, como percebido nos relatos das entrevistas, são utilidades aparentemente distantes e incomuns quando comparado ao domínio em que são produzidos, tendo assim seus usos expandidos para áreas menos óbvias. Isto aponta novamente o investigador como o agente que detém informações essenciais e poder de decisão (Swan & Brown, 2005) que repercutem para além da partilha, com informações que podem acompanhar os dados nesse sentido. Portanto, o uso dos dados de forma interdisciplinar ou pela sociedade, precisa de informação representativa para que os dados sejam objetos informacionais relevantes (CCSDS, 2012) para a reutilização.

Na primeira perspectiva, o foco está no investigador, como detentor da informação que pode direcionar onde (re)utilizar os dados de investigação. Nesta segunda perspectiva, a ênfase está na reutilização dos dados em áreas e finalidades distintas das originárias. Nas respostas aqui apresentadas, pode-se perceber a relevância da possibilidade da reutilização dos dados em domínios diferentes no qual foram produzidos, como para serem utilizados como fontes de medicamentos, para minimização de risco de inundações na população costeira ou para estudos epidêmicos. A relevância é confirmada na literatura, quando esta aponta a reutilização dos dados como a oportunidade para avanços científicos e sociais em áreas comuns, ou em áreas contempladas por diferentes contextos da investigação (Bezuidenhout et al., 2017; Henning et al., 2019; UNESCO, 2021). Além disso, a

reutilização dos dados é necessária para que a ciência participe de uma sociedade mais aberta, mas também é um dos princípios centrais da ciência aberta, sendo ainda um objetivo consequente da RDM (Bécard et al., 2016; Gezelter, 2009; Wallis et al., 2013).

Porém, para que a reutilização dos dados aconteça de forma pertinente, o resultado da análise da literatura aponta que, uma partilha adequada, bem como as práticas essenciais para a RDM, são realizadas e direcionadas para esse objetivo, de modo a oferecer suporte para que o investigador possa ser direcionado a gerir os dados para a reutilização. No sentido aqui apresentado, o investigador deve trabalhar em prol da reutilização, inclusive, se possível, sugerir utilizações presentes e futuras para os dados, sendo necessário a percepção de que existe uma organização a apoiá-lo. O contrário, pode significar mais uma sugestão de atribuição para o investigador em meio a tantas outras, correndo-se o risco de que os muitos elementos que envolvem uma RDM para a partilha, possam até resultar no depósito dos dados em meios digitais, mas sem a garantia do reaproveitamento para uso posterior, ou seja, os dados são partilhados, mas continuam invisíveis para a ciência e para a sociedade.

Na extensão das respostas para a finalidade da pergunta anterior, além da utilidade, os investigadores foram questionados quanto à percepção sobre o valor que os dados produzidos na investigação possuem em relação ao tempo de trabalho. As respostas convergiram para mostrar o valor histórico, mas também foram enfatizados outros aspectos, como o político-social ou de *marketing* institucional, mas sobretudo, os valores financeiros, estratégicos ou de mercado foram os que mais permearam as respostas.

O valor para a historicidade dos dados foram assim descritas: “(...) [é] uma coleção de dados de 1926, a mais antiga do mundo (...) (de forma) contínua (...), com observações diárias do sol (...), só os nossos colegas de Paris é que tem uma coleção similar e com quem trabalhamos em parceria” (E4). Destacou-se também que o valor pode não estar necessariamente relacionado com a sua antiguidade: “O exemplar mais antigo (...) é de 1779, mas acredito que haja mais antigo (...). No entanto, a informação que se tira deste exemplar pode ser tão interessante quanto ao que se tira de um colhido ontem” (E1). Também foi percebido o valor político-social na seguinte afirmação: “(...) [na] possibilidade dos dados virem a formar políticas públicas, é a articulação com o setor público” (E3). E ainda, o valor na divulgação da própria universidade: “Nós (...), centro de investigação

(...)[e] Universidade de Coimbra (...), temos um acervo de dados científicos que coloca de uma maneira muito particular na cena internacional (...), neste mundo do publicitar as coisas que se faz (...), há aqui uma mais valia” (E4). O valor econômico, financeiro, estratégico ou de mercado foram alvos de maior consideração, nas seguintes formas transcritas abaixo:

Quando se trabalha sobre questões da segurança social (...), qualquer política ou recomendação de política pública da área (...) tem um impacto econômico grande, seja no Estado, seja na relação na fronteira entre o Estado e o privado, ou (...) [nas] questões de saúde, acesso aos cuidados de saúde, por causa dos fluxos de saúde (...) pode haver (...) [um] valor de mercado da informação (E2).

Esses dados podem ter um valor estratégico muito grande se for apanhado pelos concorrentes, podem permitir perceber modelos de funcionamento (...), falhas que podem ser exploradas (...). Inclusive (...) um dos parceiros não tinha acesso a esses dados (...), nem a companhia aérea queria que eles tivessem (...), nem eles queriam ter acesso (...) para não se poder dizer no futuro que os produtos deles tinham se beneficiado dessa informação e eles ficarem sujeitos ao pagamento de *royalties* ou algo do gênero (...). Quando estamos a falar de empresas (...) [elas] percebem claramente o valor dos seus dados e tem receio de expor segredos do seu negócio (E6).

Isso da genealogia das pessoas, das ligações familiares, é um grande negócio (...), um dos maiores negócios do universo *Web* é a genealogia, porque as pessoas muitas vezes estão dispostas a pagar para conhecer suas origens (...). Nós já fomos assaltados por robôs (...) do Google e de alguns programas daquelas multinacionais da genealogia que queriam recolher as informações (E7).

Nós vamos ter que pensar sobre isso no futuro que é (...). todo esse trabalho que a gente fez (...), em teoria, (...) é para fins não comerciais (...) [mas] até que ponto nós estamos interessados em que as pessoas peguem os nossos dados e usem num livro e vendam? (E8).

Ao analisar o sentido da valorização apresentada nas transcrições, pode-se inferir que os dados têm um valor considerado de imediato, como os obtidos das aeronaves e oriundos da genealogia, onde há esforços de terceiros interessados em obtê-los para fins comerciais, enquanto outros adquirem valor com o passar do tempo, como os dados sobre o sol e os materiais biológicos. Em todo o caso, relata-se que o valor que os dados possuem na atualidade e o que deles se pode extrair em relação à antiguidade, são igualmente importantes. Uma possível forma de visualização dessas situações é na perspectiva de práticas necessárias para a proteção legal, técnica e de preservação, em termos de segurança, garantia da integridade, salvaguarda dos dados e armazenamento confiável com proteções físicas e lógicas (OCDE, 2021; Sayão & Sales, 2016). São práticas de proteção e preservação que repercutem positivamente na RDM como um todo, desde os investigadores terem a confiança e a segurança para partilhar os dados, quanto na acessibilidade da retenção de dados em longo prazo (Carlson, 2014; Higgins, 2008), ou seja, abrangem questões sociotécnicas em práticas da RDM, repercutindo também para a reutilização dos dados.

As respostas ainda apontam para a necessidade de uma RDM que ponha em relevância o valor dos dados. Tal relevância também pode ser evidenciada por intermédio de práticas de proteção e preservação que possam garantir o conhecimento e a utilização deste valor na atualidade e em longo prazo, sob o entendimento de que o custo da perda de tais dados, históricos, comerciais, ou de outros tipos que não estão de modo explícito nas respostas, podem, em certos casos, serem irreparáveis no sentido científico, social ou econômico. Cabe então construir uma estrutura de atenção para que o valor dos dados, independente do seu aspecto – histórico, econômico, financeiro, estratégico ou de mercado -, seja considerado com a devida importância, principalmente quando se observa a frase do sexto entrevistado, que participa de investigações em parceria com diversas empresas comerciais, a frase pode ser vista no sentido antagônico em relação à universidade: “Quando estamos a falar de empresas (...) [elas] percebem claramente o valor dos seus dados” (E6).

A sensação decorrente é que as universidades, de forma geral, estão a perder recursos irrecuperáveis, uma vez que as possíveis perdas são de difícil mensuração, seja na percepção dos prejuízos pela falta de exploração e reaproveitamento dos dados ou pela ausência de conhecimento sobre os dados produzidos em seu âmbito, o que resulta em uma lacuna imperceptível em detrimento dos benefícios sócio científicos. Assim, é necessário conhecer os dados e o seu valor através de práticas que os identifiquem, estruturas que os façam conhecidos publicamente dentro e fora da universidade, observando os cuidados com a proteção e a preservação para o uso atual e futuro. Nesse sentido, mesmo que a utilidade dos dados seja indicada pelo investigador, ainda não é possível saber todo o potencial valorativo que os dados possuem. No entanto, um esforço para perceber o valor atual dos dados e talvez o que ainda possam possuir, devem estar presentes na RDM aqui objetivada.

Os investigadores também foram questionados a respeito da existência de dados caracterizados como sensíveis e/ou reservados. As respostas apontam a disponibilização pública como um fator de risco a ser considerado: “(...) nós podemos estar a disponibilizar a localização de espécies que são muito interessantes, por exemplo para viveiristas” (E1). Mas também, tais dados, geram a impossibilidade da comunicação científica: “Existem algumas prestações de serviços em que nós entregamos relatórios (...) que não podemos utilizar essas informações para outras publicações” (E2). Ao mesmo tempo, apontam

possíveis soluções para que os dados estejam disponíveis publicamente: “(...) temos que estabelecer algum filtro em termos de dizer onde elas [as espécies] estão” (E1), ou: “temos que garantir toda a anonimização da informação quando trabalhamos com processos” (E2). Outros entrevistados apresentaram as suas opiniões, preocupações e quais práticas adotam em relação à questão da reserva e a sensibilidade dos dados, conforme as citações seguintes:

[Os] investigadores (...) submetem seus protocolos de investigação de acordo com formulários que existem (...), se estiverem abrangidos com questões de alguma sensibilidade, por exemplo (...) projetos que trabalhem com populações vulneráveis (...) ou com informação confidencial (...), devem passar pela comissão de ética (E2).

Há práticas bem instaladas, seja em armazenamento em condições de acesso restrito, protegidos com *passwords* e até proteção física, armários de acesso restrito. Dentro das próprias equipas da investigação há muitas vezes o cuidado para não partilhar os dados brutos entre diferentes elementos da equipa (...). Há uma série de mecanismos (...) que garantem que estes dados tenham o mínimo de partilha possível (E3).

Há dados sensíveis (...), no caso do lixo cósmico (...) que vem orbitando a terra e que eventualmente possa cair (...). São bases de dados internacionais, com muito sigilo governamental em que nós enquanto investigadores (...) temos muita dificuldade em [aceder] alguns dados (...), mas precisamos deles para trabalhar. (E4).

Obviamente que este tipo de informação é altamente sensível para as companhias aéreas (...), elas não querem que (...) seja de conhecimento dos seus competidores (...). É preciso um conjunto grande de garantias sobre a utilização dos dados, tipicamente há acordos de confidencialidade (...) e muitas vezes há partes dos dados que são anonimizados (...). Muitas vezes tem que pedir (...) autorização antes de fornecer esses dados para a investigação, a menos que isso tenha sido contemplado nos acordos iniciais de recolha dos dados (...). Especialmente em empresas europeias com a regulamentação geral da proteção dos dados (...), quando os dados são coletados deve ser dito claramente para o que eles serão utilizados (E6).

Nós temos toda essa preocupação no jogo (...), as pessoas têm que autorizar (...), em termos de proteção de dados dos respondentes (...), é anônimo, confidencial e tudo mais (...). (Para) fazer o *login* (...), [o usuário] cria o nome (...) e a própria senha (...), esses (...) casos nós não podemos distribuir, porque nós temos de alguma maneira uma relação com a identidade (E7).

Todos os dados que eu estou a tratar estão publicamente disponíveis na Internet (...) e os dados são puramente académicos (...). Quando a informação que foi produzida para não ser pública, é tornada pública, acho que isso deve ser feito com muito cuidado, em regra geral não deve ser feito (E11).

A sensibilidade dos dados pode limitar a possibilidade de partilha (Lane & Shur, 2010). No entanto, as respostas direcionam para que os acordos de confidencialidade, autorizações, formulários, garantias e protocolos sejam estabelecidos de modo a permitir o sigilo, a anonimização e a confidencialidade dos dados. Além disso, outros instrumentos como mecanismos de cuidados na partilha interna e as restrições de acesso, ou em termos de armazenamento, proteções físicas e lógicas e mesmo a razoabilidade, são citados como importantes para garantir a segurança e a confidencialidade dos dados. Estas medidas abrangem tanto a proteção das pessoas, governos, interesses comerciais ou cuidados com a

exposições de vulnerabilidades. Nos relatos também se observa as preocupações e o esforço buscado para encontrar soluções, de modo que não se partilhem os dados indevidamente.

Como percebido nas respostas, é requerido dos investigadores que atentem com instrumentos necessários para identificar o nível de sensibilidade dos dados (Sayão & Sales, 2020). Ou seja, perceber quais são os dados pessoais, confidenciais ou reservados, que possam ser utilizados para identificar um indivíduo, espécie, objeto, processo ou local que introduz um risco de discriminação, dano ou atenção indesejada (Eynden, 2015). Em conjunto com esta identificação, os instrumentos éticos; legais; ferramentais tecnológicos, juntamente com políticas bem definidas e ajustadas entre si, devem ser utilizadas em relação aos dados desta natureza. Esta combinação pode ser tão determinante, quanto complexa de se implementar, pois devem prever de forma geral, uma série de detalhes e características que garantam à proteção dos dados e ao mesmo tempo a partilha dentro dos parâmetros do acesso aberto.

Para minimizar a complexidade, verifica-se que a análise da literatura confirma a necessidade dos cuidados relatados, quando é salientado que a partilha adequada deve ser realizada sempre que possível. Ou seja, este cuidado deve permear a RDM com a utilização de práticas essenciais, que apontam a necessidade de constituir uma proteção legal e técnica para os dados, de modo a evitar uma partilha inadequada e proporcionar sua proteção. Os cuidados, responsabilidades, mas também a dúvida fazem parte do contexto, de modo que as influências cognitivas, contextuais e sociotécnicas descritas na definição de “uma partilha adequada” indicam a possibilidade de gerar no investigador conflitos que interferem na própria partilha, como visto na exposição do primeiro entrevistado, quando da possibilidade de “estar a disponibilizar a localização de espécies” (E1). Os conflitos podem decorrer da dinâmica entre o desconhecimento, a necessidade da partilha, e a confiabilidade da proteção e segurança para os dados sejam partilhados (Veiga, 2017).

Portanto, a proteção dos dados sensíveis ou reservados devem ser realizadas de modo claro, juntamente com instrumentos que possam ser utilizados em apoio ao investigador e de acordo com as especificidades dos dados. Por exemplo, os entrevistados citaram: o lixo cósmico; as populações vulneráveis; os interesses comerciais; e, os dados pessoais, como eletivos de sensibilidade. São abordagens diferentes entre si, onde é preciso analisar e determinar de modo nítido o que é público e o que é privado, o que é pessoal,

confidencial ou sensível, de acordo com o contexto. Sem tal determinação pode-se incorrer em extremos, quer na partilha total dos dados sem critérios bem definidos, ou mesmo na decisão de não partilhar, por receio de estar expondo publicamente o que não seja devido. Mesmo os dados sensíveis e confidenciais podem ser partilhados de forma ética e legal, como no exemplo sobre o lixo cósmico, que, apesar de sensíveis, podem ser acessados por investigadores, obviamente com a segurança e os critérios necessários. Estas são barreiras compreensíveis que tendem a não levar aos extremos da partilha. Assim, para proteger os dados neste contexto, a RDM deve possibilitar que uma partilha indevida seja evitada e permitir a segurança para partilhar os dados necessários.

6.3. RECURSOS PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO

O segundo bloco de perguntas obtém informações sobre os atores e as práticas para a gestão dos dados. A finalidade principal é identificar os atores envolvidos que podem articular e influenciar na RDM, representados pelas pessoas, pelos processos, fluxos ou divisões de trabalhos existentes e pelas ferramentas e serviços, bem como as políticas, normas e orientações adotadas para o apoio da atividade, além da colaboração e participação das comunidades envolvidas e de partes interessadas, de dentro ou de fora da universidade.

6.3.1. PESSOAS ENVOLVIDAS

Indagados sobre a existência de pessoas com conhecimentos, envolvidas ou com responsabilidades na RDM da investigação, os respondentes destacaram indivíduos com perfis informáticos, em funções informais e participantes diretos da equipe da investigação. Encontraram-se também pessoas com responsabilidades institucionalizadas em apoio à RDM. Os detalhes apresentados encontram-se nas citações que seguem.

[O responsável] é uma pessoa multidisciplinar e que realmente tem muito valor (...), porque (...) tem o perfil do usuário e o perfil do informático (...) e esta combinação é perfeita (...), por causa desta flexibilidade e desta interdisciplinaridade as coisas aceleram de uma forma fantástica (E1).

Agora com o [gestor de dados] (...), que faz a articulação direta com a equipa de investigação e traz também a sua leitura (...), nos permitiu começar a trabalhar estas questões (...). [No] nosso organograma aparece (...) não como uma função autónoma, é uma competência funcional (...), alguém que nos ajude a gerir a publicação (...), o tratamento desses dados e que apoie as equipas de investigações a garantir o *open access* e (...) os dados de investigação também (E3).

Não, (...) em projetos nunca tive nada a esse respeito (...). Ninguém responsável, não houve nenhuma formalização dessas decisões (...). Os [dados] que chegavam da análise já eram limpos

por aquele investigador que estava responsável (...). Normalmente o que acontece é que tem um investigador que limpa o “lixo” (...) e depois de limpo, partilhava com todos (E5).

Aqui a vantagem é que o criador (...) desse sistema tinha essas duas valências (...), é historiador, mas também tem profundos conhecimentos de programação e trabalho com bases de dados. Nunca houve (...) dissonância entre o historiador e o informático (...), o que as vezes não acontece nos âmbitos dos projetos (...). Nós, neste projeto que estamos a arrancar, temos uma pessoa de informática (...), que nós contratamos especialmente porque tem formação em arqueologia (...), optamos por pessoas que sejam humanistas digitais (...), que navega nos dois mundos (E7).

Nós temos o linguista computacional que tem esse acesso e que pode fazer isso (E8).

Temos que garantir que fique sempre alguém responsável pelos dados (...) quando fazemos (...) protocolos (...) [com] a comissão de ética do hospital (...), tem que haver algum de nós que fica responsável, por isso há sempre alguém que fica responsável por armazenar os dados (...), mas não é a única atribuição (...), tem que participar da investigação (...), tem que ser voluntário para colher os dados (...) e depois para fazer a gestão dos dados (E9).

Sob as respostas concedidas, percebe-se que as atividades dos investigadores em relação à RDM, seja determinada pela força de algum tipo de protocolo ou realizada de modo espontâneo, confunde-se com o processo de investigação. Assim, os mesmos investigadores que “limpam o lixo” dos dados, também recolhem, armazenam e participam das atividades básicas da investigação. No entanto, a resposta que descreve as atribuições funcionais e de como se comporta um “gestor de dados” em uma UI&D (E3), aponta para um tipo de função formalizada para apoio aos investigadores. A falta de pessoas para atribuições específicas também é percebida na declaração dos respondentes, quando se destaca a “flexibilidade” e as valências inter e multidisciplinares dos indivíduos, ou seja, enfatiza-se as pessoas que “navegue nos dois mundos”, como no caso do “linguista computacional” e dos “humanistas digitais”, de modo geral com perfis informáticos.

As respostas apresentadas podem, por um lado, representar o que diversos autores citam como questões impeditivas para a partilha dos dados (Piwowar, Day & Fridsma, 2007; Tenopir et al., 2015; Zhu, 2020). São questões que exigem dos investigadores habilidades, características e perfis técnicos para muito além da própria formação, o que parece não ser uma realidade factível para a atualidade das investigações em uma universidade. A procura por suprir tais valências pode resultar em um aumento da carga atributiva de funções para os investigadores realizarem a RDM, somado às atividades já rotineiras do processo de investigação. Por outro lado, dois relatos apontam pessoas externas à investigação como apoio à RDM, indicando uma possibilidade para minimizar tal carga atributiva. Para que isso seja viável é preciso identificar as áreas de responsabilidade (Whyte & Tedds, 2011), definir

atribuições para as pessoas participantes da gestão, de modo formal, e, se possível, como uma “competência organizacional”, pois poderia assim minimizar a questão da transitoriedade e das mudanças de acordo com decisões informais, resultando em uma forma de normalizar o apoio para o investigador na RDM que, em última instância, é quem lida diretamente com os dados de investigação e com os vários aspectos para a sua gestão e partilha.

Embora o resultado da análise da literatura aponte o investigador como ator principal, indicando que este ator deve ter a preeminência em vários aspectos das práticas para gerir e partilhar os dados, o mesmo resultado também indica que a articulação dos atores para a realização das práticas essenciais para a RDM, tem como um dos seus objetivos não causar justaposição de trabalho. Nesse sentido, cabe ao investigador participar juntamente com outros atores, por exemplo, do planeamento dos fluxos de trabalho e das atribuições das pessoas envolvidas na RDM. Assim, as pessoas e as responsabilidades na RDM devem estar divididas, sob uma estratégia que leve em consideração o contexto ambiental para que não haja sobrecargas no trabalho do investigador para RDM.

No entanto, diante das respostas aqui obtidas, cabe ponderar se o motivo da valorização dos perfis informáticos, também pode ser decorrente da falta de apoio neste sentido, somado à importância que a informatização tem para a RDM e para a partilha dos dados. Isto porque as práticas para a partilha dos dados na forma do acesso aberto requer necessariamente o uso de tecnologias, serviços e ferramentas digitais. Assim, a estruturação da RDM para a partilha dos dados, quando enfatizada às pessoas envolvidas, deve ser direcionada a um planeamento e divisão do trabalho que permitam soluções adequadas à realidade da investigação e das UI&D, de modo a evitar a justaposição de responsabilidades (e-IRG, 2016), mas também observado o apoio para as práticas atuais das tecnologias informáticas e correlatas, ou outros recursos pertinentes na RDM para a partilha dos dados.

6.3.2. PROCESSOS E FLUXOS DE TRABALHO

Quando questionados se havia um processo estabelecido ou um fluxo de trabalho para a organizar os dados geridos na investigação, os entrevistados que responderam positivamente, citaram uma série de práticas que constituem o processo ou o fluxo existente, conforme a figura abaixo e as explicações transcritas posteriormente:

Figura 36

Itens constituintes do processo ou fluxo de trabalho, em ordem aleatória



(...) quando colhemos o material, o material é processado, é preparado, é montado, é informatizado e digitalizado e só depois que entra na coleção (...), as ofertas só entram após estarem montadas, informatizadas e digitalizadas (...). Até 2017 nós tínhamos uma metodologia em que a pessoa que ia buscar os exemplares, fotografava e fazia a transcrição para a base de dados (...), acabei por desenvolver (...) outra metodologia (...) para diminuir a interação do utilizador com o exemplar, para reduzir os erros também, os enganos que aconteciam (E1).

(...) o gabinete de gestão de projetos estará a montante do processo que (...) [parte] da conceção do que é que nós fazemos com os dados que vamos recolher? (...) centra-se mais no processo de recolha e armazenamento (...) e tratamento (...) dos dados (...). [O gestor dos dados] intervém (...) depois do projeto terminado (...) [para identificar] o que é que fazemos com os dados, como é que os podemos deixar em acesso aberto, acessíveis para outros colegas da instituição. E articulamos estes dois serviços para garantir que acompanhamos o ciclo todo da gestão dos dados. (...). Nas ciências sociais (...) uma boa parte dos procedimentos que agora são tidos como procedimentos padrão na gestão de dados, se decorrem da própria metodologia de investigação e, portanto, a maior parte desses procedimentos já estavam interiorizados como (...) boas práticas de investigação, boas práticas metodológicas (E3).

A gestão de dados acontece em muitas etapas, desde o próprio levantamento da literatura à organização da informação que se recolhe aí, e como nós gerimos essa informação, até depois na aplicação de qualquer instrumento que acaba por levar a uma recolha de dados primária (...). O projeto tem os dois tipos de levantamento de dados (...), com base em entrevistas (...) feitas com profissionais de saúde, essencialmente (...) transcritas na íntegra (...), feita uma análise temática, modo misto, (...) extraídos os temas principais e as questões principais (...), fizemos (...) uma interface com determinadas características (...) virada para o estatístico (E5).

O processo era (...), as equipas que precisavam dos dados estavam claramente identificadas, essas equipas explicavam quais que eram os detalhes que precisavam dos dados e (...) depois quando o avião aterra (...), os dados coletados automaticamente (...) são transferidos para os servidores da companhia aérea (...). A companhia aérea (...), depois de obter as devidas autorizações internas da

gestão, selecionou a amostra, anonimizou algumas partes e enviou para (...) o projeto de investigação (...), para a equipa que precisava (E6).

Normalmente nos nossos projetos nós temos (...) pessoas que fazem a anotação e que introduzem os dados, que os importam para a base de dados, e depois temos os administradores que tomam decisão no nível seguinte (...), decisões já com base científica (...) que permitem a consolidação (...) e a transformação (...) desses dados (...) no produto final (...). No contexto de trabalho, normalmente há essa divisão de tarefas (E7).

Anotamos todos os registos paroquiais com uma equipa (...) de investigadores (...), recolhemos os dados do século XVI até 1800 mais ou menos, permitindo a reconstrução da comunidade a partir dessa análise formal (...). Implicava ter que olhar para as fontes, transcrevê-las e utilizar o sistema de anotação da informação (...), transformando em linguagem XML (...), [a linguagem] permitia a importação para um sistema (...) de base de dados (...). Portanto não se trata apenas de digitalizar (...) e disponibilizar, há aqui todo um processo de intervenção, tomada de decisões, de validação (...), já estamos a falar em algo que está entre o dado e a criação científica (E7).

Nós fazemos a (...) extração de dados desse *corpus* e depois nós damos todos os dados para um programador, que é a pessoa que vai colocar todos esses dados no jogo (E8).

Nós acabamos por seguir uma série de passos que em nossa cabeça já estão definidos (...). Para iniciar o processo de aquisição dos dados (...), precisamos sempre fazer em colaboração com (...) o hospital. O segundo passo é fazer um *template* (...) em que nós descrevemos aquilo que queremos fazer (...) e temos que submeter (...) à comissão de ética (...), depois o estudo começa em função do que nós planeamos e enviamos à comissão de ética (E9).

(...) numa (...) fase (...) [a] análise, elaboração, produção de interpretações (...), [permite] trabalhar com diferentes perspetivas sobre os dados (...). Sempre [houve] essa ideologia própria do método científico (...), uma distinção clara entre aquilo que observou e aquilo que se interpretou (...) e diferentes pessoas devem poder reinterpretar os dados doutra maneira e validar ou invalidar (...) isso é a base do processo científico (E11).

É possível perceber nas explanações que há uma variabilidade de etapas e procedimentos nomeados para organizar os dados geridos na investigação. A figura 36 é um exemplo de como pode haver muitas práticas designadas para a RDM. Parte dessas práticas são verificadas como fruto da investigação e da experiência dos investigadores, resultando na construção de metodologias próprias, desenvolvidas e adaptadas para a realidade, a partir do processo de investigação e observadas em frases do tipo: “acabamos por seguir uma série de passos que em nossa cabeça já estão definidos” (E9); “(...) acabei por desenvolver (...) outra metodologia” (E1); e, “decorrem da própria metodologia de investigação” (E3).

A versatilidade também é percebida nas citações que mostram práticas peculiares, como, por exemplo: “A companhia aérea (...) selecionou a amostra (...) e enviou (...) para a equipa que precisava” (E6), ou ainda, (...) “utilizar o sistema de anotação da informação (...), transformando em linguagem XML” (E7). Percebe-se também que atores externos à investigação têm papel influente no processo e no fluxo de trabalho: por exemplo, a coleta de dados realizada por terceiros, como nos casos da companhia aérea e do hospital. E ainda

na participação do gabinete de gestão de projetos, da comissão de ética e na atuação do programador ou da figura do gestor de dados.

Todas as práticas e características de trabalho juntas e não ordenadas podem tornar-se um tanto quanto confusas e imbricadas. No entanto, os fluxos e processos são ferramentas necessárias, para a organização do trabalho e na obtenção uma visão geral da RDM para a partilha dos dados, tendo em vista a percepção deste ambiente como um arcabouço variado de práticas aparentemente sobrepostas (Carlson, 2014; Corti et al., 2014; Tammaro, Sposito & Casarosa, 2019). Portanto, tal ambiente é acompanhado de desafios para formalizar um fluxo que tem por objetivo arranjar a forma de trabalhar, tendo por base as práticas para a RDM (Strasser, 2012), pois sem a clareza do que seja, ou quais sejam as práticas para a RDM, os fluxos tendem a se fundir em práticas realizadas de formas indistintas (Borgman, 2010; Lord & Macdonald, 2003; Weidner & Alemneh, 2013). Assim, a falta do entendimento do que envolve a RDM ou de processos formais que esclareçam e organizem os fluxos em práticas minimamente comuns, pode culminar em etapas tão heterogêneas que se torna inviável uma RDM proposta para a universidade.

Gerir os dados com práticas comuns não significa padronizar práticas de investigação distintas, pois, como percebido nas transcrições, o processo deve ser adaptado de acordo com a necessidade e perfil da investigação. No entanto, gerir os dados nessa perspectiva significa conceber o mínimo de práticas, que possibilitem a gestão e a partilha dos dados dentro de um fluxo organizado na investigação. A referida adaptação é ratificada pela análise teórica na definição das práticas essenciais para a RDM, quando tal definição além de indicar um fluxo e quais sejam as práticas mínimas, acrescenta que estas devem ser associadas ao ambiente. Portanto, é necessário um fluxo para a RDM que seja flexível e adaptável à investigação, permitindo a inserção de novos elementos determinantes no processo de organização do trabalho, partilha e reprodução do próprio fluxo (Borgman, 2012).

Ainda em relação ao fluxo de trabalho que envolve a gestão dos dados dentro do processo da investigação, os entrevistados descreveram as dificuldades, preocupações e problemas que provém da implantação de tais fluxos com a integração de pessoas e de práticas e das possíveis consequências, além do que pensam sobre o assunto. Estas descrições seguem nos registros abaixo:

(...) como chegar a um equilíbrio entre o que é uma exigência procedimental (...) de uma boa prática de investigação, sem passar para o lado do exagero do mecanismo de controlo que transforme a investigação em um processo burocrático estandardizado? (...). O risco dessa projetificação da ciência com base em financiamentos (...), é que não há tempo para que a ciência atinja o patamar de inovação (...) que pode atingir (...), um dos meus receios é que estes processos (...) [com o] controlo rígido, possam chegar a um ponto excessivo que limita (...) uma investigação de qualidade (...) e ficamos apenas com os investigadores mais rotinizados com burocracia (E2).

(...) a maior dificuldade é obter os dados propriamente dito (...), mesmo em um projeto europeu em que (...) as universidades são financiadas, [onde] há objetivos contratualizados com a comissão europeia (...), as empresas são reticentes a entregar os dados (...), querem garantir-se caso haja (...) uma fuga de dados (...). [Temos] grandes dificuldades (...) [para] chegar a esses repositórios técnicos (...), as empresas querem fornecer o mínimo de dados possível, ou que seja representativo (E6).

Nós discutimos (...) como integrar o pessoal da parte computacional com as pessoas que lidam com a parte aplicada da teoria, como fazer os dois se entenderem (...). Desde o início a gente foi fazendo uma construção conjunta, todos lemos juntos a bibliografia (...), nós todos fomos aprendendo juntos, aliás o *crowdsourcing* é uma coisa nova para todos nós (...), é por isso que eu acho que a gente conseguiu chegar a essa forma conjunta de trabalhar, mesmo sendo de áreas diferentes, porque todo mundo está aprendendo junto como se faz aquilo (E8).

(...) não precisamos recolher os dados, precisamos que nos deem acesso aos dados (...), [são] dados que já foram recolhidos. (...) Depois (...), tentar por sermos nós a definir como é que os dados devem ser recolhidos (...) e aí define-se um protocolo (E9).

A minha perspetiva é que, isto do *Research Data Management*, se calhar estamos a abordá-los como se fosse uma questão [de construção mais antiga] (...), não é se estão superestruturados numa coisa superformalizada com um plano guardado numa hierarquia ontológica (...), mas devíamos abordá-los como um bazar, o que significa (...) estimular os processos colaborativos. (...) Tudo que implica em ciência aberta, implica em um investimento de tempo do investigador nesse objetivo (...) para ser aberta tem que ser uma coisa muito intrínseca da própria natureza daquilo que se faz (...), se criamos a ilusão de que ciência aberta é preencher um formulário e um registo e fazer um plano, e fazer os documentos, independentemente de acharmos que é razoável ou não as pessoas fazerem alguma coisa com os nossos dados, acho que isso vai ser mais uma burocracia académica. (...) Todas as universidades têm que ter essa política estruturada (...), mas deve ser *light*, é fazer formação, é criar competências locais (...) um esforço transversal em promover esse tipo de coisa exige investimento, exige tempo, exige convencer os investigadores, e só são motivados se eles mesmos forem convencidos que alguém vai utilizar seus dados de uma maneira útil, porque senão é mais uma burocracia (E11).

No relato de dois respondentes (E6 e E9), observa-se as dificuldades na recolha dos dados para a RDM dentro do processo de investigação, são dados que apresentam valor comercial ou sensibilidade. Tais dificuldades podem levar a restrições no fluxo de trabalho, pondo em relevo a necessidade de no processo atentar para os cuidados com os dados a serem partilhados. Outra preocupação está posta em como integrar pessoas com perfis diferentes no mesmo processo, aqui expostos como a integração do “pessoal da parte computacional com as pessoas que lidam com a parte aplicada da teoria” (E8), resultando na conclusão por parte do entrevistado de que: “(...) se conseguiu chegar a essa forma conjunta de trabalhar mesmo sendo de áreas diferentes, porque todo mundo está aprendendo

junto como se faz aquilo”. Este último exemplo pode ser um indicativo da característica da articulação, principalmente porque ao observar o propósito da investigação na qual o respondente faz parte, constata-se a peculiaridade em se ter investigadores de cinco línguas diferentes realizando um projeto em conjunto na área da linguística. Tanto as dificuldades levantadas na coleta, quanto a falta de integração colaborativa podem levar a restrições no processo, o que implica a possibilidade de perdas informacionais no fluxo de trabalho, mas também em perdas por falta de ajustes entre os elementos humanos. Assim, deve-se tratar os fluxos de trabalho para a RDM de forma atenta ao processo de investigação, bem como ao comportamento dos participantes para que sejam realizados com o mínimo de perdas.

Destaca-se no relato do segundo entrevistado, o uso de termos como “exigência procedimental”, “exagero do mecanismo de controlo”, “processo burocrático standardizado”, “projetificação da ciência” e “investigadores mais rotinizados com burocracia”. Tais termos expõem o receio de que não haja tempo para que a ciência atinja o patamar de inovação, tendo como consequência uma investigação de qualidade limitada, devido aos fluxos e processos exigidos com base nos financiamentos da ciência. A questão proposta é: como chegar a um equilíbrio entre o que seja uma boa prática de investigação e essas exigências procedimentais? A questão levantada, em parte, assemelha-se ao desafio exposto por Beall (2012), em levar à consistência um modelo procedimental envolvido em uma teia de formalidades que exponenciam o aparecimento de fluxos rígidos e previamente impostos, resultantes da ávida produção para fins de recompensa acadêmica, responsáveis por levar os investigadores a procurar tempos de submissão e publicação mais curtos. A situação exposta, contrastada com um processo para a RDM que necessita de tempo e reflexão para ser realizada de forma devida, ratifica a necessidade de fluxos flexíveis, com o mínimo de dificuldades e adaptados ao ambiente, com a participação do investigador.

Outra resposta a ser destacada é a do entrevistado E11, no tocante à abordagem que se faz da construção para a RDM, uma vez que a sua opinião converge para que se tenha estímulos, em contraposição a processos “superestruturados numa coisa superformalizada com um plano guardado numa hierarquia ontológica”. Pois, segundo o entrevistado, este trabalho dentro de uma filosofia aberta: “implica em um investimento de tempo do investigador (...), tem que ser uma coisa muito intrínseca da própria natureza daquilo que se

faz” e não apenas “preencher um formulário e um registo e fazer um plano, e fazer os documentos”, porque, segundo o entrevistado, não promover um processo mais leve, formativo e com a criação de competências, resulta em mais uma “burocracia acadêmica”. Essa opinião é interessante no contexto processual porque pode indicar uma forma diferente de abordagem dos fluxos e processos para construção do modelo aqui proposto, pois o entrevistado utiliza como pano de fundo a analogia para a construção do “código aberto”. Esta analogia é encontrada no livro de Raymond (1999), *A Catedral e o Bazar*, que, entre outras coisas, aborda a criação das comunidades de desenvolvimento através de uma construção “horizontal” e colaborativa. Essa perspectiva pode ir ao encontro do que indicou a análise da literatura, a qual tem no acesso aberto seu nexó teórico. Isto significa, que um processo ou fluxo de trabalho sob a filosofia aberta, que deve abarcar uma articulação de atores, práticas essenciais para a gestão e a partilha adequada dos dados, pode ser pensado no mesmo sentido de construção comunitária e colaborativa para a RDM na universidade.

E por fim, na extensão das respostas sobre os processos e fluxos de trabalho, os investigadores também citaram as práticas de descarte e preservação dos dados. Neste sentido, assim afirma o sexto entrevistado: “Não sei, mas imagino que haja uma expectativa da companhia aérea de quando os dados não forem mais necessários para o projeto, eles sejam apagados” (E6). Esta questão remete necessariamente ao planeamento para saber o que será feito com os dados geridos após a investigação. No mesmo contexto foram citados também os riscos de perdas de dados, inferindo sobre a necessidade de preservação:

(...) dados (...) em papel estão lá (...), e lá a gente não sabe o que é que lá tem, isto é um problema grave (...), [estes dados] não estão tratados, não estão eventualmente catalogados (...), são 150 anos de dados (...), este legado de grandes séries de dados é extremamente importante ainda mais hoje (...). No caso [dos dados] do sol as coisas estão melhor (...), a partir dos anos 2000 (...) se fez a recolha digital (E4).

Esta é uma razão por que queremos que (...) os dados fiquem hospedados na Universidade de Coimbra (...), porque meu contrato é temporário (...), eu não sei se meu contrato vai ser renovado daqui a três anos, não sei se vai haver financiamento novamente (...), portanto (...) os dados vão ficar salvaguardados (E8).

(...) nem sempre existe uma consciência clara do que pode acontecer, quando uma tecnologia muda, uma plataforma deixa de estar disponível, as pessoas que criaram o projeto já não estão ou porque morreram ou porque foram trabalhar em outros contextos e muitas vezes não há essa garantia (...). Muito da história desses arquivos digitais é uma história em certo sentido, de arquivos mortos (...), muitos deles não (...) estão adaptados para serem responsivos (E10).

Esses planos tecnológicos de preservação de formatos (...), quanto mais estruturado for a informação, se não for num formato reciclável (...), sempre que sai do formato de texto começa a

ser muito complicado garantir (...). Havia um tempo que toda gente estava a fazer *backup* dos seus dados de investigação para CDs (...) hoje não se usa mais CDs (...). Acho que o *GitHub* vai estar cá daqui [a um longo tempo] (...) seguramente, porque há tantos *softwares* lá, tantos milhões de pessoas, tanto a partilhar (...), mas estão lá também porque os formatos são, formatos de texto partilháveis (...). Eu morro amanhã e alguém pode continuar, mas para isso (...) é preciso que esteja não só lá os dados, esteja o código fonte (...) e a documentação necessária para as pessoas entenderem (E11).

A necessidade da preservação dos dados já foi relativamente abordada aqui. No entanto, este é um tópico de bastante interesse e de discussão fundamental para a RDM objetivada na partilha dos dados em acesso aberto, por isso pode ser explorado por muitos vieses. Este assunto é citado de forma recorrente pelos investigadores porque a preservação é uma prática que envolve a relação direta para se evitar a perda dos dados. Nos trechos transcritos acima, apresenta-se os riscos da perda dos dados em papel proporcionado pela falta de digitalização, os dados citados embora de valor reconhecido, permanecem neste sentido ainda desconhecidos. A transitoriedade das pessoas e a obsolescência tecnológica também foram pontuados como itens a serem observados para a preservação, evitando as perdas dos dados ou que estes sejam ofuscados pelo tempo e pelo desconhecimento.

O acesso aos dados no futuro requer que eles sejam preservados hoje, mantendo suas características. Para tanto, as práticas devem ser realizadas em um processo contínuo de gestão que permeia todo o ciclo de vida da investigação, normalmente com o início em um planeamento dos dados que serão obtidos, mas também envolve questões a serem dirimidas tais como: quem é o responsável pela preservação, que tipos de controle (sociais, éticos e legais), em qual infraestrutura, quais são os dados e com quais recursos financeiros será realizada a preservação dos dados (Sayão & Sales, 2013, 2015). Tais questões, bem como a necessidade de preservação aqui exposta, podem ter seu direcionamento nos fluxos e processos de trabalho, isto porque quando observado o resultado da análise da literatura, tanto as práticas essenciais, quanto uma partilha adequada dos dados e a articulação dos atores, são constituídas por fluxos e processos que possibilitam estabelecer e assegurar uma organização apropriada para preservar os dados, que incluem o planeamento, a documentação, a atenção com a sua proteção, as ferramentas digitais, as atribuições, políticas e recursos estruturais, entre outros atores e práticas diretamente ou indiretamente relacionados com a questão da preservação, mas também relacionados ao processo e fluxos de trabalho para gerir e preservar os dados de investigação.

6.3.3. COLABORAÇÕES E INFLUÊNCIAS

As questões concernentes às colaborações, foram realizadas para a percepção da influência de atores da RDM no ambiente da investigação, direcionados sob duas perguntas: a primeira questiona os investigadores se há o costume de trabalhar de forma colaborativa com instituições internas ou externas à universidade, ou com profissionais da área da informação? A segunda pergunta é sobre quais atores poderiam ser úteis na colaboração com a gestão dos dados para a investigação? As respostas indicam que as colaborações influenciadoras vêm e podem vir de vários quadrantes da ciência e da sociedade.

A figura 37 apresenta os principais atores influenciadores citados, agrupados em três tipos, são eles: 1) atores internos; 2) atores externos; e, 3) atores transversais, os dois primeiros em relação à universidade, o terceiro pode conter em si atores a influenciar tanto de forma interna quanto externa. E ainda, a tabela 21 elenca os tipos de colaboração realizados por esses atores. As respostas dos entrevistados, a contextualização e os detalhes das explicações que formaram a figura e a tabela, seguem abaixo e após elas.

Figura 37

Atores internos, externos ou transversais que podem influenciar à RDM

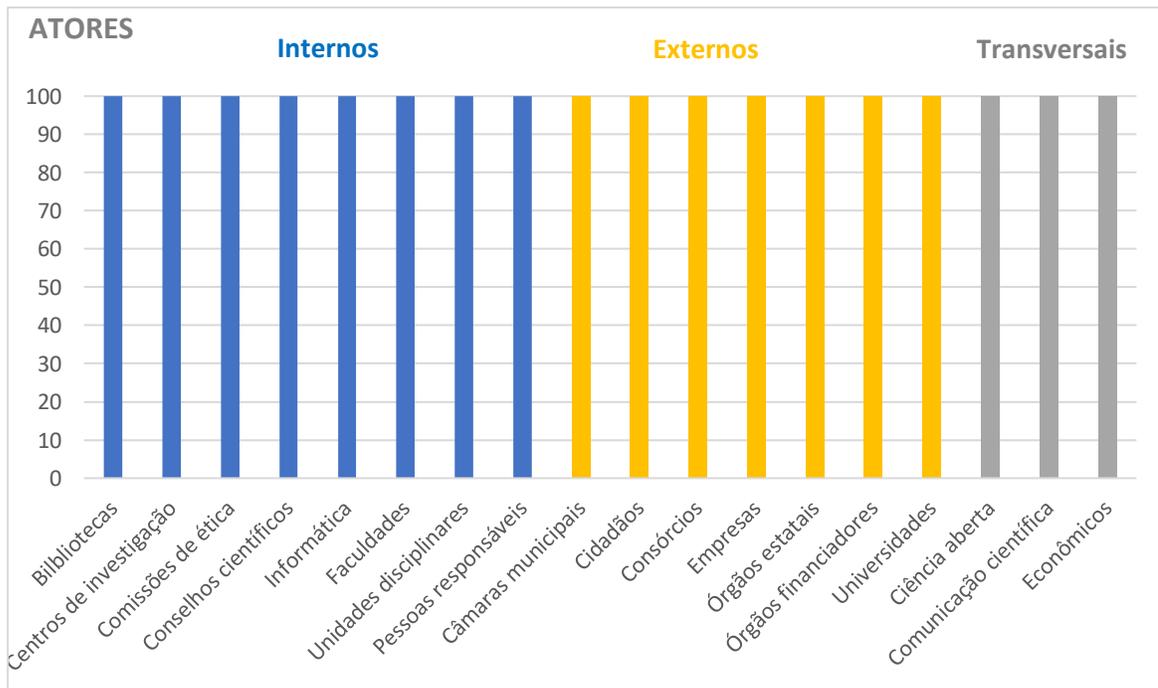


Tabela 21

Tipos de colaboração dos atores

Tipo de colaboração dos atores identificados nas transcrições	
Acordos científicos interdisciplinares	Protocolos com municípios
Articulação com centros de investigação e empresas	Decisões com órgãos internos da UI&D
Consórcios internacionais para a partilha dos dados	Acordos comerciais com órgãos estatais
Consultoria informal no ambiente da UC	Colaboração da sociedade com participação cidadã
Parceria com empresas e universidades	Publicação em revistas para comunicação científica
Financiamento por agências nacionais e internacionais	Parcerias disciplinares entre faculdades da UC

Quando queremos fazer qualquer coisa diferente consultamos o [nome da pessoa] (...) do departamento de Engenharia Informática (E1). (...) Nós temos o projeto de ciência cidadã em que o cidadão colabora (E1). (...) Há uma colaboração histórica entre nós e a faculdade de farmácia (E1). (...) Nós tivemos projetos financiados (...) dos Estados Unidos (...) e financiados também pela FCT (E1). (...) Para que o trabalho seja publicado, as revistas científicas exigem o depósito de exemplares [biológicos] (...) e que seja indicado (...) a localização, que exemplares são e onde estão depositados para estudos futuros (E1).

(...) temos aqui algumas parcerias e estamos a fazer candidaturas que implica também parcerias com engenharias (...) [e] na área da justiça (E2). Temos parceria (...) [com] o PASSDA⁴⁹ (...), está ligada a uma rede europeia (...) que é o CESSDA⁵⁰ (...), onde já existem práticas de partilha da informação (...). É uma mudança cultural muito grande (...), a informação era retida ao máximo (...) principalmente nas ciências sociais e humanas, hoje em dia as coisas mudaram muito e é importante que (...) trabalhe em equipa, em rede (...) [isto] começa cada vez mais a ser uma prática corrente em projetos de consórcio, ou seja, que envolve várias instituições a volta de uma temática (...), é uma realidade (...) estimulada também ao nível da união europeia em que privilegia em seus financiamentos que haja colaboração entre áreas científicas (E2). (...) Essa já é uma realidade em curso (...), o trabalho colaborativo e articulado com centros de investigação e empresas (E2).

Há protocolos firmados (...) nos dados geofísicos (...) com o pessoal da meteorologia, da geologia (...), temos acordos firmados com o (...) Instituto Português do Mar e Atmosfera (...), temos alguns investigadores que estão a trabalhar com algumas empresas da aeronáutica, com a ANA também dos aeroportos (...), a ciência se faz em rede, articulando esforços e saberes (E4).

(...) [no âmbito da] direção e conselho científico é (...) que discutimos modelos (...), as boas práticas (...). Tivemos uma discussão interessante: se podemos tratar a história de vida como dados, se podemos utilizar a expressão dados para uma metodologia que é qualitativa (E3). [A] comissão de ética (...) entre outras atribuições, tem também (...) [a] verificação da adequação dos procedimentos da investigação (...), [a comissão] tem feito um caminho (...) de sistematização desses procedimentos e está nesse momento a fechar uma avaliação das principais questões, que inclui também questões de proteção (...), de recolha, armazenamento e tratamento dos dados (E3).

No consórcio (...) tinha vários parceiros de vários países, empresas e universidades (E6).

No quadro europeu, um projeto que envolveu várias universidades (...) [Também em outros] dois projetos (...), um com a câmara de Óbidos e outro com a câmara de Ílhavo, protocolados com a

⁴⁹ Production and Archive of Social Science Data - consórcio para infraestrutura portuguesa de recolha, arquivo e disseminação de dados sobre atitudes, valores e comportamentos sócio-políticos.

⁵⁰ Consortium of European Social Science Data Archives - fornece serviços de dados em larga escala, integrados e sustentáveis para as ciências sociais.

faculdade de letras e outra com o CESXX (E7). (...) Esse sistema vai ser infraestrutura para um projeto financiado pela FCT (...), em âmbitos de tese e doutoramento (E7).

(...) são pessoas que gostam de alguma forma de língua (...), os nossos participantes na Holanda (...) têm uma *newsletter* e um grupo muito forte de pessoas da sociedade em geral (...) [que] gosta de participar dos eventos promovidos por eles (E8).

Sempre tem (colaboração com) outras instituições [universitárias]: italianas, espanholas, francesas e alemãs (E9).

Há muitos utilizadores (...), criamos também alguns sistemas colaborativos (...), quando alguém (...) vem olhar para uma base de dados desta e começa a seguir, deteta erros (...) e portanto, o [nome da investigação] começou a incorporar estas partes mais cooperativas (E11).

Os atores citados neste agrupamento de respostas, bem como os seus tipos de colaboração e influências, abrangem agentes que são encontrados sob esta mesma ênfase na literatura, percebidas através das agências de fomento (Borgman, 2012; Shearer, 2015), representadas nas transcrições pelas agências nacionais e internacionais. Mas também no âmbito da comunicação da ciência, enfatizado pelas transformações neste assunto (Autran & Borges, 2014; Borges, 2006), como as exigências das revistas, que fazem parte do sistema de atores influentes e que repercutem na RDM (Freitas & Leite, 2019). Ou ainda, na colaboração da sociedade, com a participação de uma ciência cidadã (Paul Ayris et al., 2018).

A ciência, de um modo geral, e especificamente a ciência aberta, pressupõe a colaboração como um dos seus pilares (Fecher & Friesike, 2014). Essa ciência colaborativa, referenciada na *e-Science*, envolve um número abrangente de características associadas à forma com que a investigação é conduzida em um ambiente articulado, quando da perspectiva de colaboração (Jankowski, 2007), como nos casos aqui expressos, ou seja, através de acordos científicos interdisciplinares, da articulação entre os centros de investigação, das parcerias entre as faculdades da UC e com outras universidades. Mas também, são vários os consórcios internacionais que se organizam sob a filosofia da ciência aberta e da partilha dos dados (UNESCO, 2021). Assim, sob a perspectiva de parcerias e colaboração, são formalizados acordos comerciais com empresas e órgãos públicos. No entanto, iniciativas informais como consultorias pessoais, internas e de forma esporádica, além da influência direta de outros órgãos da universidade, são destacadas nas transcrições.

Ao observar nas respostas, as dissimilaridades e a abrangência dos atores que efetivamente colaboram e influenciam na RDM do ambiente, verifica-se a dificuldade para que a partilha dos dados seja realizada em acesso aberto pelos muitos atores existentes e

peculiaridades influenciadoras. Ao mesmo tempo, se pode entender que sem cooperação entre atores articulados, torna-se inviável que essa partilha seja realizada de modo adequado, a partir do entendimento de que os atores precisam trabalhar em conjunto para que a RDM possa obter uma compreensão comum para o desenvolvimento de soluções colaborativas na universidade (Shearer, 2015). Assim, as práticas e a articulação de atores para a RDM, deve canalizar os esforços para que tais colaborações tenham sua utilização organizada e direcionada de forma positiva para a gestão e partilha dos dados.

Na segunda pergunta, os entrevistados responderam quais atores, de acordo com a sua percepção, ajudariam sendo capazes de interagir para colaborar de forma importante com a gestão dos dados nas investigações. As respostas mostraram a necessidade de apoio sem um padrão muito específico, mas ressaltando áreas importantes de suporte para a RDM, como pode ser visto na tabela abaixo e nas declarações que seguem.

Tabela 22

Atores que poderiam ser úteis para a colaboração com a RDM

Atores úteis	Motivos citados para a utilidade
Biblioteca	Pelas características dos padrões, classificações, metadados e taxonomias.
Comissão de ética	Para apoiar as diversas áreas de conhecimento da universidade.
Gestor de ciência	Realizar parcerias internas e externas, políticas organizacionais, intercâmbios dos centros, centralizar o conhecimento e apoiar o investigador.
Informáticos	Interagir com softwares, servidores, plataformas, gerir os dados e infraestruturas que exigem este tipo de acompanhamento.
Pessoas dedicadas	Garantir a privacidade, guarda e acesso aos dados, questões éticas e aspectos legais.
Jurista	Apoio técnico especializado na área.
Gestor de inovação	Apoio técnico especializado na área.

Eu diria que (...) as bibliotecas, (pois) já têm (...) padrões (...) [e] formatos (...) que acabam por servir de inspiração a esse mundo das coleções, mas do ponto de vista institucional nunca tivemos uma colaboração direta (E1).

(...) dentro da Universidade de Coimbra pressionamos muito para que houvesse uma comissão de ética ao nível da universidade que pudesse trabalhar com as diferentes áreas do conhecimento (...), porque na área das ciências sociais e humanas ainda não havia uma resposta institucional da universidade, que pudessem resolver os problemas que se exigem hoje no financiamento da investigação (E2).

(...) nos quadros de apoio técnico, um jurista (...) [e um] gestor de inovação (E3).

Um bom gestor de ciência que nos ajude (...) articular uma série de coisas que agora estamos a falar (...), até (...) parcerias e centralizar (...) uma série de conhecimentos e uma série de políticas organizacionais e intercâmbios dos centros com os outros centros, é importante e não temos (...). Teria que ser alguém que ajudasse os investigadores do centro em suas parcerias internas e na sua

organização interna e comunicação com o exterior (...), teria que ser alguém com estabilidade (...). Agora não temos, mas já tivemos uma pessoa dedicada (...) à plataforma, é difícil (...), agora somos nós (...), tínhamos um informático que também (...) foi embora, também saiu (...) e para gerir esses dados todos (...) que produzimos, essa interação toda com servidores aqui e acolá e *softwares* (...), poderia ajudar bastante (E4).

(...) de facto em qualquer projeto com ajuda de alguém que possa resolver problemas na infraestrutura é relevante (...). Pessoas que percebessem de base de dados (...), por vezes há problemas de infraestrutura que exigem acompanhamento informático (E7).

Nós gostaríamos de ter no nosso centro, [uma pessoa] que fosse só responsável por essa parte administrativa barra gestão, que daria apoio técnico para questões (...) [de] acesso a servidor (E8).

Se houvesse uma pessoa dedicada (...) [para] garantir que os dados (...) privados [fossem] guardados em um local que não é acessível (...); questões de ética (...) [e] todos os aspetos legais (...), eu diria sim (...), faz todo o sentido ter (...), eu e os meus colegas não teríamos (...) que aprender para fazer dez coisas diferentes (...) e aí acho que sim, teríamos vantagens (...). Outro nível de processamento (...) [seria] anonimizar os dados (...), [verificar os] valores que estão em falta (...), bases que não foram adquiridas. (...) Antes de começar a usar os dados, aquilo que chamamos de pré-processamento, existe um trabalho enorme (...) para que eles comecem a ser usados (...) aí, não, cabe as equipas de investigação (...) Dentro das nossas equipas de investigação, não temos essa especificidade, ter uma pessoa que faça isso, outra pessoa que faça aquilo, isso (...) é para equipas grandes (E9).

O bibliotecário, porque [o nome da investigação] tem outra dimensão que é como nós utilizamos um sistema de classificação dos textos através das taxonomias, ou seja, [nós] colocamos os próprios utilizadores a criar metadados através do modo como vão lendo os textos e vão atribuindo categoria aos textos, o [nome da investigação] joga um pouco com (...) a dimensão classificatória, a dimensão de construção de categorias, que depois tornam a informação pesquisável, partilhável etc (E10).

Os aspectos importantes para a RDM a serem destacados nas entrevistas, estão nos atores que podem colaborar de forma generalizada, ou seja, são gestores, juristas, informáticos, bibliotecários que também estão referenciados na literatura (Jones, Pryor e Whyte, 2013) e que têm potencial para exercer suas atividades de modo abrangente na universidade. Também as respostas fazem referência a um indivíduo que possa exercer atribuições diretamente ligadas à RDM para realizar articulações de forma interna e externa. Esses e outros atores são vistos tanto no campo empírico quanto na literatura, como um apoio necessário para que os investigadores não estejam sobrecarregados e conseqüentemente, ocorra uma partilha inadequada ou mesmo a não partilha dos dados.

Embora sejam importantes tais colaborações, principalmente as que sugerem uma articulação de abrangência alargada a nível da universidade para repercutirem adequadamente na RDM e na partilha dos dados, exige-se também um empenho a ser realizado pela gestão da própria universidade, a fim de que os diversos tipos de recursos sejam viabilizados. Por exemplo, a articulação da área da biblioteca em apoio às investigações, por si só necessitaria de um grande esforço em vários aspectos gerenciais

para que tal articulação fosse viável, ou ainda, a necessidade de pessoas estáveis funcionalmente para a realização do trabalho, dependem diretamente da influência e colaboração de outros atores para que esta necessidade, desta forma, possa ser cumprida. Assim, pode-se inferir que é necessário o estabelecimento de uma rede colaborativa, com articulações abrangentes que permitam melhor explorar o potencial comum, pois as articulações em iniciativas isoladas dentro das UI&D ou das investigações, embora sejam pertinentes, sem uma visão macro para a RDM, podem continuar a invisibilizar os dados ou apenas utilizar parte do seu potencial.

Em geral e pelas características observadas, uma articulação colaborativa e abrangente para a RDM, deve ter a participação dos gestores e da liderança acadêmica da universidade (Borgman & Bourne, 2022; Freitas & Leite, 2019). Além disso, também se apresenta a necessidade de ter atores humanos para auxiliar de modo relevante os investigadores em questões que envolvam a RDM no ambiente, respeitando a abrangência necessária e conforme as necessidades da própria investigação. São responsabilidades que podem estar divididas com especialistas que sejam capazes de fazê-lo devido à sua formação técnica para tal, por vezes com mais propriedade que o próprio investigador.

O resultado da análise da literatura para uma partilha adequada dos dados e para a articulação dos atores para a RDM, confirma a necessidade do estabelecimento de um conjunto de atores, que permite incluir tanto os gestores e a liderança acadêmica da universidade que estejam associados e sistematizados. Assim, os atores descritos, além de outros que sejam pertinentes para gerir e partilhar os dados conforme proposto, devem estar organizados de tal modo que a necessidade de colaboração para a RDM e para a partilha sejam supridas de forma sistematizada.

Além da necessidade da colaboração para a RDM, foram realizadas objeções em relação aos limites colaborativos dos profissionais que poderiam ajudar na RDM:

A minha opinião é que há muito da informação do contexto, que profissionais que não tenham estado envolvidos no processo não vão compreender, e por isso até pode ser arriscado passar-lhes esta informação e deixá-los trabalhar esta informação, se a informação for assim passada sem mais nada. Portanto, a menos que a finalidade tenha sido mesmo essa de partilhar (...), eu acho que os dados devem ser reservados a quem está envolvido diretamente no projeto e sabe aquilo que está a fazer e como é que se fez, como princípios (E5).

No nosso tipo de investigação eu diria que não [necessitaria] (...), o tipo de análise que se faz sobre esses dados são análises altamente técnicas que tipicamente são feitas por *data scientist* (...), a olhar para os dados e ver quais ferramentas (...) podem usar (E6).

Acho que cada atividade é tão específica que é difícil tentar fazer as coisas de uma forma genérica (...), têm (...) determinadas formas de agir (...), ia ser uma confusão e ninguém ia se entender (...). Aceito que as coisas possam ser centralizadas (...) se houver dimensões para isso, de haver pessoas dedicadas (E9).

As considerações no sentido de limitar ou mesmo de não existir a colaboração, pode ser percebida, em certo sentido, por uma falta de esclarecimento a respeito do tema, como, por exemplo, até mesmo o modo de fazer ou de entender a pergunta sobre essa questão, pode ter levado a tais respostas por parte dos investigadores. Portanto, a educação, a promoção e o treinamento na chamada educação aberta, ou educação em ciência aberta, pode ser um fator importante de modo a comunicar corretamente, a identificar as necessidades de treinamento dos investigadores, suprir lacunas no conhecimento e habilidades em relação ao envolvimento com a ciência aberta (O'Carroll et al., 2017; Ayris et al., 2018), mas não só, essa prática é relacionada também com as especificidades da RDM, com o esclarecimento do que significa uma partilha adequada dos dados, com a adaptação a novos processos e tecnologias oriundos de uma filosofia aberta, entre diversos itens a serem organizados neste sentido.

Em outras respostas, observa-se também os atores e o contexto da colaboração proporcionada por práticas disciplinares já estabelecidas e por exigências de órgãos de fomento direcionadas pela ciência aberta. Assim, pode-se perceber com mais clareza as considerações dos investigadores e as possíveis consequências oriundas das colaborações no ambiente da investigação. São atores, culturas e métodos com influências relevantes que influenciam no ambiente da investigação e da RDM, conforme segue:

(...) os botânicos têm (...) estruturas muito mais bem oleadas e a funcionar há muito mais tempo do que qualquer outra área da biodiversidade (...), essas redes (...) de cooperação para investigação são redes muito antigas que hoje se modernizaram com as novas tecnologias (E1).

(...) foi um processo que chegou (...), sentimos essa pressão (...) e foi essa necessidade de responder aos financiadores (...) que sentimos a necessidade de facto de ter uma comissão de ética que trabalha, olhe para os projetos, para os protocolos de investigação e valide, corrija, ajude a melhorar as práticas de investigação na instituição (E2). (...) Os projetos da FCT embora estimulem, estão organizados de uma forma disciplinar (...) e limita essa efetiva colaboração. Nesse mesmo sentido as exigências e consequências decorrentes da ciência aberta que afetam o ambiente das UI&D (...) são exigências que vêm de cima para baixo, muitas delas justificáveis, mas não vem o apoio financeiro necessário para que elas possam operacionalizar facilmente (E2).

(...) antes falávamos em investigação e metodologia, agora falamos em acesso aberto, gestão de dados (...), isso tem levado a pensar em alguns procedimentos institucionais de forma uniformizada (...), leva que sejam criadas também outras carreiras técnicas, o gestor da ciência, o gestor de dados, o especialista das questões éticas (...), há um conjunto significativo de investigadoras e investigadores (...) diante (da) (...) falta de oportunidades para fazer investigação, que estão a mobilizar-se para este tipo de carreira (...), o que é interessante (...) porque garante uma articulação entre os dois mundos e que permite de facto aliviar quem faz a investigação destas tarefas mais burocráticas (...), isto exige recursos que as instituições não têm, exige que haja uma estrutura técnico-científica alargada, que quanto mais cresce esta estrutura de recursos humanos, maior é a pressão para obter financiamento e o financiamento é cada vez mais difícil (...), há aqui uma tensão grande que obriga as instituições a ponderarem bem se vão dar ou não vão dar este passo (E3).

(...) devolver à sociedade o conhecimento (...), isso precisa ser encarado (...) e é legítimo. Quem está a sustentar o centro e os outros centros e as universidades é a sociedade com seus impostos (...), é preciso se informar as pessoas o que é que se faz, a importância (E4).

No passado valia tudo, entre aspas (...), isto é relativamente recente, também agora, penso que nos projetos europeus (...) [ao] financiar um projeto, todos os artigos e todas as publicações (...) têm que ser de domínio público (...) coisa que no passado não existia (...), era publicado numa revista e a revista tinha que ser paga (E9).

Na minha perceção, de uma pessoa que está nisso há muito tempo (...), da importância de colaborar, partilhar (...) as coisas nesses últimos cinco anos deram um salto absolutamente inacreditável, e este salto é fruto dessa nova ideologia (...), a ideologia também da *open Science*, mas sobretudo a ideologia do *Code For All, Everyone Can Code*, e todo o impulso daquilo que é *Data Science* (...), uma nova área do saber que é provocada pela acumulação maciça de dados (...) que eu acho que as ferramentas todas, vão ter um impacto (...) [colaborativo] no conceito de *Open Science* (E11).

A primeira influência colaborativa observada nas narrativas, a da cultura da investigação ou tradição disciplinar (Thessen & Patterson, 2011), pode ser vista aqui de modo positivo na área da botânica. Neste caso e com este tipo de influência, parece ser mais fácil a obtenção de resultados favoráveis em relação à partilha dos dados, mas isso não exime dos problemas que impedem a mesma partilha, pois a necessidade de adaptação e de “modernização com as novas tecnologias” (E1) podem trazer consigo também barreiras e conflitos, ou mesmo a tendência a não estabelecer limite para a partilha. Assim, a atenção para as investigações onde se tenha uma boa tradição para a partilha, deve ter a mesma atenção dada a áreas que não as tenha, o que pode mudar é o aspecto de cuidado a ser observado.

A segunda influência colaborativa observada está na necessidade da adaptação da investigação às transformações oriundas da ciência aberta e do acesso aberto, mais especificamente sobre as práticas exigidas pelas agências financiadoras (Borgman, 2018). Assim, o segundo entrevistado destaca que: “foi um processo que chegou (...), sentimos essa pressão (...) e foi essa necessidade de responder aos financiadores” (E2). Este tipo de influência pela pressão, exercida através de instrumentos exigidos pelos financiadores para

a abertura dos dados, aparentemente afeta a investigação no aspecto da colaboração de modo também positivo, ou seja, para que haja um planejamento para gestão e partilha dos dados, com o uso adequado da RDM no ambiente das universidades (Shearer, 2015), mas também pode ser percebida como uma colaboração geradora de tensões e conflitos, causadas pela própria resistência à obrigatoriedade da partilha (Veiga, 2017).

É importante ainda, atentar aos comentários direcionados às adaptações e transformações da ciência na atualidade, em frases como: “devolver à sociedade o conhecimento” (E4) e “no passado valia tudo”. Estas frases em seus contextos podem ser vistas como uma internalização oriunda da transformação e das propostas da ciência aberta para a investigação atual. O entrevistado número onze, indica que as transformações ocorridas sob a filosofia “aberta”, com suas ferramentas a evoluir, tende a mudar inclusive a própria ciência aberta em termos de possibilidades colaborativas. Isso remete ao uso de ferramentas de trabalho colaborativos *online* (Ram, 2013), onde as comunidades são criadas para facilitar a comunicação da equipe e a organização do trabalho colaborativo, lá os dados e materiais são compartilhados, desenvolvidos e reutilizados. Ainda sob a direção das mudanças e transformações, na perspectiva do terceiro entrevistado, há uma tensão na qual as exigências da ciência aberta geram novas carreiras e oportunidades: “o que é interessante”, mas também requer novas estruturas e financiamentos: “o que é cada vez mais difícil” (E3).

No contexto colaborativo exposto, as influências e as dualidades podem ser determinantes para a ação do investigador, ou seja, como o investigador recebe as influências da colaboração, também determina sua atitude em relação à RDM e a partilha dos dados. Nesse sentido, a colaboração e as interferências decorrentes devem ser observadas de forma a dirimir os seus aspectos negativos. Nesse sentido, o resultado da análise da literatura indica que para uma partilha adequada dos dados, o investigador deve ser observado de modo especial sob este tipo de influência. Isto porque a consciência do investigador para a partilha encontra na transversalidade da ciência aberta, na cultura da investigação ou na tradição disciplinar, aspectos positivos e negativos como fatores constituintes de atenção, para que as colaborações que influenciam na investigação sejam direcionadas em benefício da RDM e da partilha. É preciso, de igual modo, atentar para

novas formas colaborativas a surgirem no ambiente “aberto”, como novas práticas a serem adequadas ao contexto colaborativo da RDM na universidade.

O financiamento da ciência, bem como os aspectos econômicos e as consequências decorrentes, outra vez ganham relevo nas respostas dos investigadores, agora evidenciando um nível maior de influência exercida na investigação por este tipo de colaboração, como pode ser visto nas asserções abaixo:

(...) as ciências da saúde (...) avançou muito mais rapidamente porque tem um impacto econômico e há muito dinheiro envolvido (...) que depois é transposta para o mercado (...). Nas ciências sociais e humanas (...) não tem o mesmo impacto em termos econômicos que seja acessível, em termos de empresas (...). Em Portugal isso foi mais relevante a partir do momento em que muitas instituições (...) começaram a receber financiamentos internacionais (...) onde houve de facto uma maior pressão por procedimentos (...) que salvaguardem todas as questões éticas que possam estar envolvidos nos projetos (E2).

Um dos grandes problemas é sempre (...) ou o grande problema de toda gente é o dinheiro (...), claro que o dinheiro ajuda muito (...), as universidades têm uma missão científica e social muito grande (...) [mas] muitas vezes gerem (...) numa visão muito empresarial (...), agora (...), sem dinheiro não se fazem coisas (...), mas tem (...) que mudar a sua política (E4).

Esse projeto teve financiamento da ação COST⁵¹ (...), no sentido de participação em eventos (...) que acabou agora, faz um mês (...) e a minha ideia é conseguir um financiamento externo (...), mas (...) tudo que eu estou fazendo para o português (...) serão usados para o dicionário que nós estamos desenvolvendo no projeto estratégico do nosso centro (...), eu consegui casar esse projeto com o projeto da Universidade de Coimbra (...) que tem três eixos (...), aí já consegue popular a nossa base de dados, com esses exemplos que eu já vou coletando nesse projeto paralelo (E8).

Nas respostas apresentadas e em outras já transcritas anteriormente, pode ser verificado que o financiamento da investigação com recursos públicos e sob a exigência de agências nacionais ou internacionais, direcionados pela ciência aberta, é um agente de colaboração importante para a influência na gestão e partilha dos dados de investigação (Fecher & Friesike, 2014). Nesse sentido, a força econômica influi no andamento da RDM das investigações em vários aspectos (Higman & Pinfield, 2015), obviamente que a RDM e a partilha dos dados dependem deste tipo de recurso para que a sua institucionalização de forma colaborativa na universidade possa ser desenvolvida. Portanto, é necessário que estratégias sejam elaboradas para que os recursos econômicos ou mesmo a escassez deles, possa ser administrada de modo que a RDM para a partilha dos dados esteja em constante desenvolvimento.

⁵¹ A European Cooperation in Science and Technology (COST) é uma organização de financiamento para a criação de redes de investigação entre cientistas para impulsionar os avanços da investigação.

A depender do tipo de abordagem e direcionamentos para os investimentos, que os gestores universitário realizam, seja nas UI&D ou da gestão maior da universidade, pode gerar estagnação da RDM e levar à sua precariedade, bem como, no sentido inverso, pode levar ao desenvolvimento apropriado para a RDM. Assim, este tipo de ator transversal é determinante para a colaboração e para que as práticas para a RDM com atores articulados resultem em uma partilha adequada dos dados na universidade.

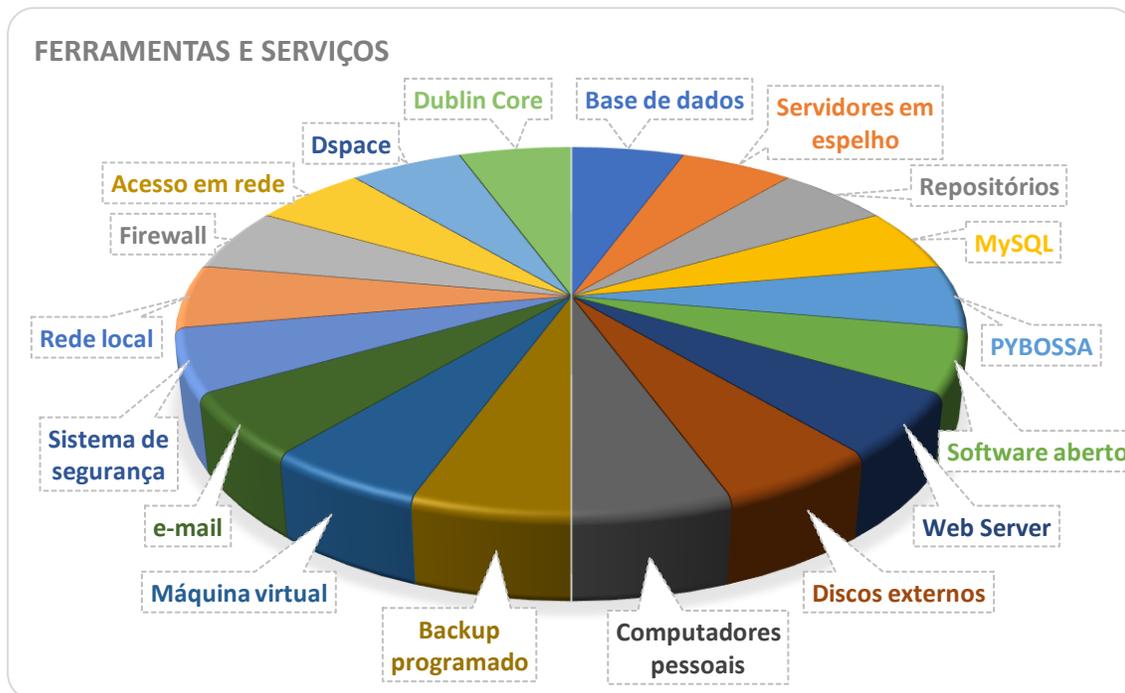
Esta gama de atores aqui referenciados, de modo geral, são facilmente identificáveis como internos ou externos à universidade. Isto é importante porque o relacionamento com esses atores para a RDM, a depender desta localização, tende a ser diferente em termos políticos, formais e de atribuições de trabalho. Foi também nomeada aqui um outro tipo de influência – a transversal -, como, por exemplo, a influência da ciência aberta, econômica e financeira e a da comunicação científica. São agrupamentos de atores que possuem em si as duas características – internas e externas -, por exemplo, dentre os vários atores da comunicação científica existem atores internos e atores externos que influenciam na RDM, o mesmo pode ser pensado para a ciência aberta e sobre as questões financeiras. Embora os atores que neles estejam contidos sejam os mesmos que são tratados individualmente como internos e externos, pela força desses agrupamentos, existe um poder influenciador próprio que deve ser observado e lidado em forma de colaboração para a partilha dos dados e as articulações e práticas para a RDM na universidade.

6.3.4. FERRAMENTAS E SERVIÇOS

Ao serem perguntados sobre a existência e uso de ferramentas ou serviços para a RDM, os investigadores relataram a utilização de infraestruturas institucionalizadas, equipamentos para o armazenamento e serviços para a segurança dos dados, também o uso de bases de dados, linguagens computacionais e *softwares* de código aberto, bem como os sistemas informáticos próprios ou desenvolvidos por terceiros. A figura 38 representa as ferramentas e serviços citados durante as entrevistas, outras características para este tipo de infraestrutura destinadas à RDM também foram declaradas. Após a figura seguem as respostas transcritas, assim como a continuidade das perguntas sobre o tema.

Figura 38

Ferramentas e serviços citados nas entrevistas



(...) [o material da investigação que] se consegue transformar numa versão digital é a informação contida numa imagem do exemplar (...) que pode ser vista imediatamente na imagem, ou transferida para uma base de dados onde pode ser feita uma pesquisa (...), disponível *online* (E1). (...) Nós temos o (...) [*software*] de ciência cidadã (...), nós atribuímos um grau de utilizador com base em seu histórico de contribuições (...). Os dados, apesar de algumas vezes não corresponder aos nossos padrões internos (...), não estão com o mesmo *standard* que nós conseguimos (...) [mas] nós podemos fazer um aprimoramento desses dados. (E1).

Estão em curso infraestruturas (...) nacionais, há uma gerida pelo Instituto de Ciências Sociais [de Lisboa] (...), é a disponibilização da informação resultante dos projetos de investigação, em particular bases de dados quantitativas, mas também informação qualitativa (E2).

Nós temos um gabinete de tecnologia da informação (...), temos dois servidores (...) em espelho para garantir que não se perdem dados (E3).

A base de dados, nós começamos com Access mas (...) eram muitas referências e migramos para o *MySQL*, neste momento esta é a infraestrutura de base de dados que sustenta um sistema que tem (...) um tradutor (...), uma linguagem de notação (...) que permite uma anotação rápida na informação documental histórica e sua integração na base de dados (...) [para que] possam ser alvos de cruzamentos bastante sofisticados que na [disciplina de] história são inevitáveis (...). As bases de dados (...) estão abrigadas no servidor da universidade e uma das nossas preocupações foi torná-las abertas ao público (E7).

Nós iniciamos o projeto com um experimento de *crowdsourcing* na plataforma PYBOSSA⁵² (...), mais tradicional (...) lá da Holanda (...), eles tem no servidor local deles uma instalação (...) então

⁵² PYBOSSA é um framework construído para plataformas de *crowdsourcing* ou de ciência cidadã.

todos os dados estavam lá (...). Neste momento nós temos dois alunos do departamento de Engenharia lá de Israel (...) e eles estão fazendo a implementação computacional (E8).

(...) as pessoas nos enviam (...) o modelo, “corremos” o modelo dele nos nossos dados e enviamos-lhes o modelo treinado (E9).

Na página dos agradecimentos está lá uma lista de *softwares* que nós utilizamos e esta lista é uma lista enorme (...), uma grande parte (...) é *software* de código aberto (...), algumas tecnologias que utilizamos (...) tem uma longevidade (...) significativa (...), o modelo de dados que foi desenvolvido foi específico e (...) tem uma arquitetura bastante complexa, essa informação técnica é uma informação que nós preservamos (...), meu colega está a tentar (...) garantir uma modularização mais fácil, ou seja, menos dependente dele próprio e para garantir sua longevidade (...) e não fique obsoleto (...). Neste momento ele está muito dependente ainda das pessoas que o conceberam (E10).

Nesses últimos anos aconteceram uma série de mutações tecnológicas na possibilidade de partilharmos os dados (...), uma série de convergências de tecnologias e práticas (...) desenvolvidas pela comunidade de código aberto (...), que permite fazer aquilo que já fazíamos a um nível muito mais sofisticado (...), que não implica (...) processos de instalação complicados (...) que sempre foi uma grande limitação que tínhamos, se alguém quisesse usar (...), na prática alguém tinha que estar por perto (E11). (...) A *GitHub* (...) [há um] esforço em tornar aquilo compreensível a pessoa mais vulgar (...). [O] *Jupyter Notebooks* (...), [onde] eu coloco o acesso vivo aos dados e, portanto, mostro o raciocínio ao vivo (...), eu posso ter meus dados originais (...) para permitir misturar a interpretação, com o processo dos dados originais, num único documento que é partilhável (E11). [o nome do projeto] antes era uma aplicação que se instalava, agora é uma *Library* que passa a funcionar dentro do *Jupyter Notebooks*⁵³, o código fonte da *Library* está no *GitHub* (E11).

A partir destes relatos sobre o tema, pode ser percebida que a utilização das ferramentas e serviços para o tratamento com os dados é realizada por intermédio de práticas como a digitalização do material de investigação (E1); uso de infraestruturas tecnológicas fornecidas por instituições nacionais (E2), internacionais (E8), da própria UI&D (E3) e da universidade (E7), com ferramentas e serviços associados ao uso de sistemas informáticos proprietários (E7), que podem estar abertos (E10) ou que foram desenvolvidos para o uso na investigação (E1, E7, E10). São coleções de instrumentos de apoio em amplitudes variadas para as práticas da investigação e da RDM que fazem parte da dinâmica dos investigadores. Estas coleções, por vezes identificadas também como *e-Science* ou ciberinfraestruturas (Hey & Trefethen, 2003), fazem parte das transformações vivenciadas na atualidade e que direcionam à atenção aos materiais para a RDM. Por sua amplitude, um dos desafios neste contexto é o correto dimensionamento das infraestruturas organizacionais e tecnológicas, pensadas para serem articuladas em uma RDM para a partilha dos dados sob o ambiente da universidade e em acesso aberto.

⁵³ O *Jupyter Notebooks* é um projeto de código aberto para dar suporte à ciência de dados interativa e à computação científica.

Ao observar a figura anterior e as dimensões apresentadas, percebe-se que, para realizar uma RDM que possa melhor aproveitar o potencial das ferramentas e serviços neste nível de abrangência, é necessário o uso de propriedades intrínsecas das redes colaborativas e que sejam multidisciplinares. Tais instrumentos, tecnologias e serviços para apoio da RDM podem exercer participação essencial para desenvolver e articular atores e práticas para gerir e partilhar os dados de investigação. Essa amplitude de ferramentas e serviços, direcionados aos meios digitais em um ambiente produtor de volumes e características variadas de dados digitais (Harvey, 2010), pode ser vista para a RDM no mínimo em dois aspectos também citados pelos entrevistados, a interoperabilidade e a sustentabilidade.

O aspecto da interoperabilidade foi destacado pelos entrevistados, enfatizando a necessidade da sua viabilidade, ao mesmo tempo que citam as barreiras, conforme as respostas abaixo:

Muitas das coisas que nós fazemos, fazemo-las porque só fazendo assim somos interoperáveis, conseguimos dialogar (...), eu para trabalhar preciso de informações dos outros e, portanto, eu mais que compreendo que eles precisem da minha, da minha coleção, e, portanto, esta coisa da interoperabilidade é uma coisa que nos preocupa (...) é nesse sentido que todos trabalhamos (E1).

[a questão] não foi trabalhada (...), conversamos (...) sobre isso, só que até há muito pouco tempo não houve financiamento. Tem que ver que muitos desses projetos e dessas bases de dados cresceram de iniciativas individuais (...), talvez um dia essa integração possa existir, acho que a infraestrutura é suficientemente flexível e os tipos de dados (...) para estarem integrados em outra infraestrutura (...) inclusivamente da universidade (...), [de modo a] alojar esse tipo de produto (E7).

Não, nenhuma [dificuldade para a interoperabilidade], o único entrave aqui (...) [é] pelo trabalho, pela dedicação, pelos custos que estes dados têm de os disponibilizar de uma forma gratuita (...), por exemplo, eu tenho um colega que está a tentar implementar aqui na universidade (...) uma base de dados comum (...), que é todos aqueles dados que participam de projetos que devem ser disponibilizados (...), acessíveis para qualquer um e (...) usá-los nas suas atividades de investigação (...), isso só não acontece porque não há uma boa vontade em disponibilizar os dados (...), em termos de condições técnicas (...) tudo isso é ultrapassado (E9).

(...) qualquer citação do [nome do projeto] pode ser referenciada (...), se nós quiséssemos extrair os metadados e criar uma espécie de catálogo que fosse interoperável, por exemplo com o sistema Dublin core ou com o sistema *Dspace* ou outro tipo de sistema (...). Para aumentar a interoperabilidade (...), do ponto de vista da extração dos metadados (...), teríamos que desenvolver um código específico para isso, para criar essa interface entre o sistema externo e o [nome do projeto] (...). Apesar dele ser muito dinâmico (...) esse dinamismo está dentro da sua própria bolha, (...) fechado em si mesmo (...), em todo o caso (...) temos (...) duas aplicações que estão aqui integradas, (...) portanto (...) extraem metadados que estão codificados nos textos e (...) que mostra uma certa interoperabilidade (...), escrevendo o código certo é possível ir lá buscar informação (...), a nível de indexação e de catalogação noutros sistemas (...), não é um sistema que não seja possível torná-lo interoperável com outras aplicações e com outras plataformas (E10).

Há questões de interoperabilidade dos dados, é tecnicamente possível ligar essa informação (...) com o ficheiro (...) da biblioteca nacional de Portugal e buscar as coisas que eles publicaram ao longo da vida (...). Por exemplo, uma coisa relativamente fácil de se fazer, tecnicamente, é pegar

uma ficha qualquer do estudante da Universidade de Coimbra e pesquisar automaticamente o que ele publicou no ficheiro da biblioteca nacional. O ficheiro [a ser interoperável], do ponto de vista informático (...), se você está interessado em saber todos os que vieram da Bahia entre 1750 e 1820 e estiveram na faculdade ao mesmo tempo de pessoas de Lisboa e do Porto que estudavam leis, isto não consegue, mas com [o nome do projeto] consegue (...) [para isso] transforma aquilo em ficheiros de textos (...) que tem a capacidade de meter isso no *GitHub* (E11).

Observa-se que a tradição disciplinar (E1) traz o aspecto da interoperabilidade como positiva quando se trata deste tipo de trabalho. Outra observação, é a percepção por parte dos investigadores que responderam sobre o assunto, de possuírem dados e infraestruturas suficientemente flexíveis para serem interoperáveis (E7) e de não haver dificuldades técnicas para tal (E9, E11), relatando inclusive de que forma seriam realizadas tais interoperabilidades (E10, E11). Assim, para que estas e outras possibilidades apresentadas sejam postas em prática, faz-se necessário que os recursos sejam projetados, um acompanhamento mais próximo em vários níveis seja proporcionado, de modo que a articulação com outros atores permita uma interoperabilidade organizada e comum, a partir dos recursos na universidade. Neste sentido, pode haver entraves para a realização da interoperabilidade desejada, pois há fatores específicos das disciplinas, mas também de modo geral, a nível da universidade, que podem impedir que os investigadores caminhem para o objetivo interoperável. Portanto, tais fatores devem ser observados e levados em consideração para que seja realizada uma partilha adequada dos dados.

É obvio que se os dados não forem partilhados não há interoperabilidade, conforme argumentado pelo nono entrevistado: “isso só não acontece porque não há uma boa vontade em disponibilizar os dados” (E9). Portanto, é necessário que as ferramentas e serviços sejam trabalhadas juntas com outras competências, de modo que a RDM seja realizada sob uma infraestrutura não somente ferramental e tecnológica, mas esclarecedora da importância relacional para os seus usuários. Nesse sentido, Borgman e Bourne (2022), apontam a necessidade de se trabalhar com infraestruturas de conhecimento, observada as características de um conhecimento comum, com recursos partilhados por um grupo de pessoas que está sujeito a dilemas sociais (Hess & Ostrom, 2007), sem desconsiderar que as tensões entre as partes interessadas da universidade são abundantes (Borgman, 2018).

A interoperabilidade entre as infraestruturas que produzem e gerenciam os dados, precisam ser desenvolvidas para que o aspecto colaborativo da RDM possa ser maximizado. Este aspecto é fundamental para a criação de valor nas instituições ao nível da gestão da produção científica e disseminação dos dados (Lewis et al., 2009). Embora a interoperabilidade na literatura da área esteja direcionada aos repositórios, é necessário também pensar na interoperabilidade dos dados armazenados em múltiplos sistemas (Wyborn & Lehnert, 2016). Segundo Borgman e Bourne (2022), as comunidades geralmente se agrupam em torno de repositórios com vários graus de maturidade e sustentabilidade, um problema identificado na ciência aberta (US National Science Board, 2005). Como os excertos também apontam para essas ferramentas, pode-se inferir que, sem relegar a interoperabilidade dos repositórios, a atenção também deve ser direcionada para outros sistemas e ferramentas de informação no âmbito da instituição e fora dela.

No sentido técnico, percebe-se que são abundantes as questões e todos os tipos de desafios para a interoperabilidade, desde a administração dos bens comuns, envolvendo recursos centralizados e descentralizados, até lidar com grandes silos de dados (Borgman & Bourne, 2022). Os níveis muito diferentes de atuações das infraestruturas, como o da universidade, demandam uma gestão que possa conceber diretrizes e monitorar o desenvolvimento de uma estrutura que direcione tal interoperabilidade tecnológica, mas não se limitando a esta. Assim, não se pode deixar de lado as especificidades dos dados que aqui são caracterizados como heterogêneos, atentando ainda que a preocupação com a interoperabilidade deve ser realizado no sentido *bottom-up* e *top-down*, ou seja, de modo a envolver a universidade com múltiplos olhares em níveis gerais, juntamente com as especificidades da investigação.

A sustentabilidade é o segundo aspecto a ser observado na literatura e nos relatos, onde ambos ratificam a amplitude dos dados e da infraestrutura tecnológica para a RDM na universidade. Sob tal aspecto estão reunidos uma gama importante de ferramentas e serviços para que a RDM e a partilha dos dados possam ser realizadas, são práticas e preocupações que também foram tratadas e referenciadas em respostas anteriores, remetendo a ideia do “longo prazo”. Assim, percebe-se que a sustentabilidade dos dados de investigação refere-se à sua preservação, acessibilidade e interoperabilidade para o futuro. Estes, são temas

imbricados, onde a sustentabilidade dos recursos de informação não é um problema novo, mas considerado como complicado pelo fato de que uma proporção crescente de dados se manifesta em formato digital (Lavoie, 2012).

A sustentabilidade aqui é abordada especificamente em termos de ferramentas e serviços, ou seja, para que os dados sejam sustentáveis, as ferramentas e serviços também precisam possuir este atributo. Assim, as estruturas e infraestruturas devem ser projetadas de forma a garantir o acesso aos dados de investigação para uso atual e posterior, para a salvaguarda e para a abertura, mas também devem ser verificados para que não se tornem obsoletos, de modo a prejudicar uma RDM para a partilha em acesso aberto ou inviabilizar a reutilização dos dados. Exemplificado com a expectativa positiva do investigador abaixo:

Mais uma coisa que a revolução tecnológica proporcionou (...) é que se foram criando formatos, cada vez mais fáceis de usar para dados estruturados em texto (...), primeiro em XML, depois em JSON (...) e esperamos conseguir lê-los daqui a dez, quinze anos (E11).

As práticas sustentáveis diretamente ligadas as ferramentas e serviços, quando confrontadas com o resultado da análise da literatura, juntamente com ações de governança⁵⁴, são desenvolvidas em questões como: a continuidade do desenvolvimento e manutenção da infraestrutura; a elaboração de planeamento, estratégias e documentações de modo a garantir o acesso aos dados no presente e no futuro; o monitoramento para evitar a obsolescência; e, a preservação dos materiais. São práticas que devem ser observadas para que as ferramentas e serviços sejam sustentáveis e proporcionem a sustentabilidade. No entanto, se não forem planeadas e efetivadas neste sentido, as ferramentas e serviços podem ser utilizadas, mas mesmo assim não proporcionar a sustentabilidade, como, por exemplo, no caso do armazenamento dos dados. Ou seja, mesmo os dados estando armazenados, sem as ferramentas e serviços necessários para que sejam recuperados, a sustentabilidade é inviabilizada, bem como os benefícios decorrentes da gestão e partilha em acesso aberto para a reutilização dos dados.

⁵⁴ A palavra governança já citada anteriormente por meio das referências, tem nesta tese o significado semelhante ao da governança de dados, como: "The exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets" (Brackett, & Earley, (2009), ou seja, o modo de conduzir ou regular o conjunto de processos e decisões por quem tem a responsabilidade da administração.

Os investigadores também foram questionados especificamente sobre o armazenamento, a segurança e os meios com que os dados são partilhados no âmbito da investigação. As respostas indicam que embora sejam comuns as práticas do uso dos dados por equipamentos pessoais, a maioria dos respondentes expressam cuidados com os dados através da utilização de cópias de segurança em rede institucional.

Nós temos uma máquina (...) que está numa rede local, que é a nossa base de dados mãe (...), [a coleção] é depois migrada para uma *Web Server* que é da instituição (...), no centro de informática (...). Duplicamos a base para o servidor (...) todos os dias (E1).

Estamos também a proporcionar (...), sempre que necessário (...), aos investigadores (...), condições de armazenar sua informação (...) seja em discos externos, seja em computadores pessoais com acesso codificado (E2).

Na maior parte dos casos os projetos optam por armazenar os dados em discos externos, porque isso também permite um nível de segurança física maior (...). O computador físico onde está armazenado os dados, tem ligação ao servidor (...) e há um *backup* programado (E3).

Os dados (são) armazenados localmente e depois descarregados para o (...) computador do investigador (E5). (...) Esses dados estariam alojados num local comum aos investigadores envolvidos nesse projeto e poderiam ser acedidos também pelos outros, desde que fizesse sentido no contexto daquela tarefa (...). É normal as pessoas que estão envolvidas numa tarefa, usar seus computadores pessoais ou aos que lhes estão atribuídos e descarregam ali os dados (...) porque é o local de trabalho habitual (...), depois cada um decide se quer pôr num sítio que é acessível por mais pessoas (...), é uma coisa que é decidido caso a caso (E5). Têm muitas equipas, tem muita formação das equipas, há pessoas que são muitas descuidadas com os dados de outros, há outras que não (...), eu não sei se há uma preocupação, eu tenho essa preocupação, eu ainda hoje partilhei um vídeo de uma entrevista e escrevi mais que uma frase, repetida de formas diferentes a dizer: atenção, não partilhar, não disseminar e querendo partilhar (...), tem de se pedir autorização (E5).

(...) os dados [são] coletados automaticamente (...), quando o avião aterriza são transferidos para os servidores da companhia aérea (...) [para] depois enviarem aos investigadores (...), os dados que foram enviados para os investigadores foram utilizados nos computadores pessoais (...), mas não partilhavam esses dados com ninguém (E6). (...) Eles [os investigadores] estão a trabalhar em conjunto sobre os dados (...), não há uma transferência, uma partilha (...), é um *asset* que vários investigadores daquela equipa (...) têm acesso (E6).

É da UC, isto é, uma máquina virtual que está alojada no [centro de investigação], a universidade dá-nos acesso a uma máquina virtual (...) que corre as nossas bases de dados (E7).

O jogo vai ficar armazenado na Universidade de Coimbra (...), no servidor (...), o *corpus* final nós ainda não decidimos onde que nós vamos armazenar, porque nós temos que criar algum tipo de ferramenta que permita a busca desses dados (E8).

As vezes essas colaborações já são tão antigas, nós já conhecemos as pessoas há tantos anos que a confiança é total (...) [então] nós damos os dados num e-mail (...) sem haver o mínimo problema (...), porque sempre os dados são anonimizados (E9). (...) Mantemos os dados na (...) Universidade de Coimbra, é o mais típico de acontecer (...) com todas as garantias de segurança dos dados estarem encriptados (E9). (...) As bases de dados [da epilepsia] estão num servidor que está aqui no departamento (...), a base de dados principal agora está na Alemanha (E9). (...) Outra coisa que pode acontecer é (...) [outros] que também querem ter acesso aos dados (...) [ou porque] entrou mais tarde ou [quando] o projeto já acabou (...), temos que pedir autorização a todos aqueles que participaram do projeto (...) isso acontece muitas vezes (E9).

Os dados estão neste momento numa máquina virtual num servidor da Universidade de Coimbra (...), o arquivo está programado para ser copiado duas vezes por dia, há um sistema de segurança que nós fazemos uma avaliação periodicamente (E10).

Os meios de armazenamento, segurança e os *softwares* para a utilização dos dados, são primordiais na dinâmica da investigação. Evitar a perda dos dados por falta ou precariedade no armazenamento é um dos fatores importantes na RDM, que resultam em uma atenção para as falhas de *hardware*, *software* ou mídia, na infecção por vírus ou ataque de hackers, falhas de energia, erros humanos causando deleção ou mudança nos ficheiros (Sayão & Sales, 2015). Estas são práticas que devem ser planeadas e preparadas antes do início da investigação para que os dados estejam seguros durante o projeto (Jisc, 2021) e partilhados para que possam ser utilizados e reutilizados. Nos excertos, embora a maioria das práticas incluam o *backup* em servidores da UC para garantir a segurança dos dados, também é perceptível práticas de partilha entre investigadores por meios que sejam aparentemente inseguros como e-mail e guarda em computadores pessoais, sob justificativas da existência de anonimizações e regras para que os dados sejam partilhados.

Borgman e Bourne (2022) dizem que há pouca consistência encontrada nas instituições sobre como os dados de investigação são mantidos. Os autores afirmam que algumas mantêm repositórios para seus professores e alunos, outras contam com repositórios institucionais e utilizam *pen drives* ou discos removíveis armazenados em escritórios individuais ou de posse de alunos. Nesse sentido, em meio às peculiaridades multidisciplinares e investigativas da universidade, os dados guardados apenas em equipamentos pessoais ou discos externos, motivam a necessidade de prover a segurança e possibilitar a minimização dos problemas e prejuízos decorrentes, sendo necessário uma organização para que esses mesmos dados não corram o risco das perdas informacionais.

Porém, como quase todos os aspectos que envolvem a RDM na universidade, as relações entre os atores que podem colaborar para suprir as necessidades não estão suficientemente alinhadas (Borgman & Bourne, 2022). Cabe assim, reafirmar a importância de iniciativas que envolvam um planeamento para delinear como os dados serão armazenados e apoiados durante e após a investigação, como o acesso e a segurança serão gerenciados e articulados com recursos que possibilitem a interoperabilidade e a sustentabilidade, pois se o armazenamento e a segurança dos dados não forem observados

em práticas essenciais para que a gestão seja realizada de modo pertinente, é provável que as outras iniciativas não de completem do modo esperado, resultando, conseqüentemente, em prejuízo para a reutilização dos dados sob a partilha em acesso aberto.

Os entrevistados que discorreram sobre o tema desta seção, ao mesmo tempo que expressam a possibilidade de serem interoperáveis nas suas investigações e demonstram que as ferramentas e serviços podem sustentar os dados com boas práticas de armazenamento e segurança, também manifestam preocupações em termos dos riscos e da necessidade em garantir uma infraestrutura tecnológica adequada para gestão dos dados. As preocupações são acompanhadas por dois fatores importantes, o primeiro são as condições para que tais ferramentas sejam postas em uso, a segunda está na necessidade da implantação, administração e suporte para uma infraestrutura física. Dentro dessas perspectivas, os investigadores apontaram quais seriam as dificuldades na forma abaixo:

É mais trabalho (...) e mais infraestruturas, por exemplo (...) [para] garantir (os) ativos da informação (...), uma instituição precisa ter espaço físico para armazenar a informação e espaço físico (...) de acesso público, precisa ter servidores (...) de acesso limitado (...), acessos em rede (...), *firewalls* todas bem montadas (...), a garantir que elas existam ao máximo e estejam disponíveis para os investigadores, tudo isso custa dinheiro, ter espaço físico, ter prateleiras, organizar a informação (...), serviços informáticos competentes (...) que apoie os investigadores a fazer esse processo de forma segura (...). No caso das ciências sociais e humanas, os principais riscos (...) são os computadores, onde a gente armazena a informação (E2).

(...) o investimento [para infraestrutura e recursos] tem que ser feito a nível nacional e é reforçando os orçamentos das instituições, garantindo (...) previsibilidade dos fundos disponíveis, (...) uma estabilidade dos financiamentos das instituições, garantindo (...) um financiamento de base (...) que lhes permite assegurar os seus serviços essenciais (...) para que tenha a liberdade e flexibilidade suficiente para investir (...) nestas funções que são as que as diferenciam (E3).

(...) dados (...) em papel estão (...) correndo risco de vida, porque esses dados antigos têm suporte relativamente frágil (...), não estão nas melhores condições de armazenamento (...), [para] estes dados históricos mais antigos, é preciso fazer um esforço em recuperar, porque são importantes para estudos de longa duração (...) e está a se fazer uma grande esforço em por estes dados acessíveis aos investigadores (...), em termos de criar ferramentas, *softwares*, servidores dedicados onde se põem os dados (E4).

Eu estou (...) num outro projeto, e é absolutamente irritante, sistematicamente o link de partilha da pasta não funciona, e depois não se consegue aceder aos dados que supostamente deveriam ser partilhados conosco, isto é bastante frustrante e cria bastante desmotivação (E5). Não sabemos se os problemas são das ferramentas ou se os problemas são do conhecimento das pessoas acerca dos conceitos que estão por trás das ferramentas (...), penso que são um pouco as duas coisas (...) porque também as aplicações não são claras em relação ao modo como funcionam e ao que permitem fazer (E5).

Os suportes informáticos são fundamentais (...), é impossível [sem os suportes] recolher uma diversidade (...) de informações para tentar reconstruir esses percursos de vida (...), o investigador não precisa se deslocar ao arquivo para fazer a recolha (...), o que não acontecia a dez anos atrás (...), ou ter meios para digitalizar (...), mas aqui [há] custos envolvidos (...), nós trabalhamos com

software open source (...), os custos aqui é da mão de obra (...), durante muito [tempo], isto viveu de iniciativas, ou individuais, ou então de financiamento de entidades (...), no nosso caso estamos a falar de avanços mais ou menos “caseiros” e não financiados (E7).

Os termos aqui expressos indicam alguns tipos de preocupações encontradas por estes investigadores para o uso de ferramentas e serviços na RDM, são termos como: “É mais trabalho (...), tudo isso custa dinheiro” (E2); “o investimento tem que ser feito a nível nacional (...) um financiamento de base (...) que lhes permite assegurar os seus serviços essenciais” (E3); “dados (...) em papel estão (...) correndo risco de vida (...), não estão nas melhores condições de armazenamento” (E4); “porque também as aplicações não são claras em relação ao modo como funciona” (E5); “estamos a falar de avanços mais ou menos caseiros e não financiados” (E7). Dentre essas frases destacam-se as preocupações com os recursos financeiros que possibilitam a estruturação física e tecnológica para a RDM, preocupações pertinentes, pois existem sempre custos de estrutura, primordial ao suporte geral das atividades e dos projetos (Relatório UC, 2019), o que remete para a necessidade de saber como as instituições financiarão os serviços de apoio da RDM, resultando no fato de que muitas dessas infraestruturas ainda precisam serem desenvolvidas (Katz et al., 2018).

Outra preocupação que aqui aparece como possível barreira para garantir uma infraestrutura tecnológica adequada para a RDM, está relacionada com a usabilidade das ferramentas e serviços (Veiga, 2017). Na discussão se os problemas são das ferramentas ou são do conhecimento das pessoas acerca dos conceitos que estão por trás das ferramentas, a falta de conhecimento, facilidade ou formação do investigador para o uso da tecnologia possibilita gerar tensões pela dificuldade que as ferramentas tecnológicas podem proporcionar por carências verificadas no investigador, mas também pode ser considerado nesse contexto a falta de “clareza” das ferramentas para que o uso pelos investigadores seja realizado de modo facilitado.

As respostas sobre as preocupações derivadas das ferramentas e serviços para uso na RDM, podem ser vistas também sob o ponto de vista dos investimentos, do aumento de volume de trabalho, da preservação dos dados que precisam ser digitalizados, da usabilidade, entre outros que devem ser alvos de atenção. Logo, tendo em vista a necessidade da observação das preocupações relativas ao contexto, como parte de uma possível solução, pode ser pensado envolver equipes técnicas de várias áreas e da

governança da universidade. No entanto, em todos os sentidos abordados, os problemas revelados pelos investigadores estão intercalados e necessitam também de um arranjo lógico, com questões mais amplas a serem abordadas para as estruturas a nível de universidade, ou mais específicas tal como a preservação dos dados e a usabilidade dos materiais, porém todas devem ser tratadas com igual importância, dada a interdependência que as une à RDM e ao ambiente.

Na percepção de parte dos entrevistados, as possíveis soluções para as dificuldades encontradas no contexto das ferramentas e serviços para a RDM, passam por realizar serviços integrados e interdisciplinares para a concepção da infraestrutura, armazenamento e segurança para os dados e para a gestão dos dados, como seguem nas narrativas abaixo.

Em termos tecnológicos (...), eu penso que o caminho é definir algum grau de embargo para acessar a informação (...), fornecer os dados parcialmente ou não fornecer de todo (...), o processo ainda não está implementado (...). Ou seja, nós podemos ocultar a informação acedida por texto na nossa base de dados, mas fornecemos a imagem, por exemplo se as coordenadas estiverem na etiqueta elas vão ser fornecidas também (...), há algumas tentativas de se fazer um *blur* (...) para desfocar a imagem neste ponto, para que não se veja (E1).

(...) se coloca uma questão de haver escala em termos dos centros de investigação, o que implica, melhores processos de integração (...), eu colocaria aí a gestão numa lógica de escala gerida a nível de universidade (...), para a universidade também poder, em alguns casos, salvaguardar a existência (...) das condições infraestruturais: informática; arquivo da informação; etc., com serviços partilhados (...) para que as instituições possam desenvolver todas para um nível elevado de exigência (...), porque os financiamentos são mais concentrados e há uma otimização de recursos. Havendo estruturas administrativas que sejam comuns dentro de um centro de investigação é mais fácil desenvolver essas práticas (...), ainda há um caminho longo a percorrer dentro das universidades, (...) minha preocupação é na garantia que existam centros de investigação (...) com serviços de apoio para que a responsabilidade não caia toda em cima do investigador (E2).

(...) permitir a sincronização do trabalho de vários investigadores ao mesmo tempo, porque isso em termos de gestão de projetos foi algo que se verificou que era muito necessário (...) e que esses dados todos, pudessem ser sincronizados com a base de dados principal, tem que se ter um administrador que depois valida (...). Transformar (...) em uma ferramenta de equipas (E7).

Acho que as coisas estão a se encaminhar para haver na gestão de dados científicos (...) uma democratização (...), [que] tem a ver com as ferramentas (...), porque tornou-se possível tecnicamente (...) [com] as ferramentas chamadas controlo de versão, que permite qualquer pessoa ir buscar uma versão do código, trabalhar nela, fazer as suas alterações e depois devolvê-la ao património comum (...), isso permite a colaboração de várias pessoas nos mesmos textos, mas textos estruturados (...) e a essa capacidade de duplicar os dados e o código, juntou-se também todo o processo de reporte de problemas (...), gestão de comunidades (...) e o esforço (...) em termos de tornar a tecnologia acessível (...), torna-se cada vez mais fácil as pessoas irem buscar dados (E11).

As indicações de soluções sugeridas pelos investigadores, seja no aspecto da segurança local, com necessidades específicas para a anonimização e proteção dos dados, passando pela concepção de estruturas comuns, otimização de recursos, sincronização do

trabalho e utilização de ferramentas para uso em equipe, apresentam-se, em sua maioria, como itens que apontam para soluções que envolvem algum tipo de articulação. Assim, para que uma RDM possibilite o uso de ferramentas e serviços tanto para gerir quanto para partilhar os dados heterogêneos da universidade, é preciso articular e dimensionar infraestruturas tecnológicas e organizacionais, além de determinar o grau de investimento na preservação em longo prazo, com uso e construção de políticas e estratégias voltadas para os cuidados com os dados (Sayão & Sales, 2020). A falta ou precariedade das ferramentas e serviços para que a RDM seja realizada através de uma infraestrutura tecnológica que contemple as suas necessidades, torna-se impeditiva para a realização da própria partilha (Kim & Stanton, 2016).

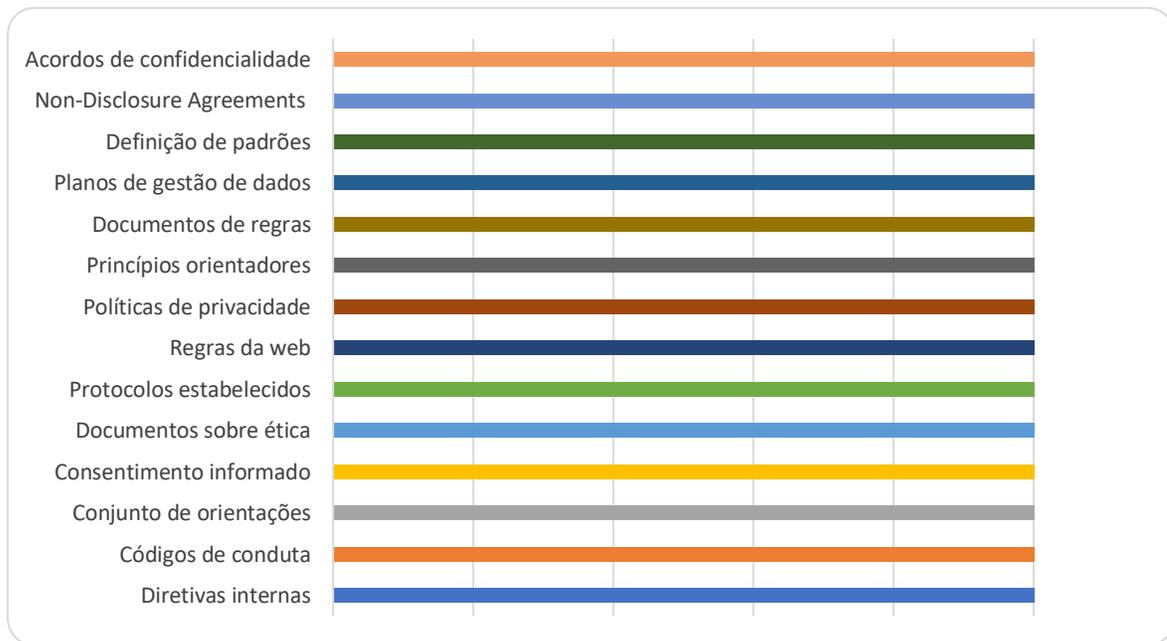
A infraestrutura tecnológica como um dos pilares da RDM (Brown, Bruce & Kernohan, 2015; Kenney & McGovern, 2005; Shearer, 2015), observada também nos outros blocos das entrevistas, precisa ser bem delineada no ambiente da universidade para permitir que várias práticas, como a guarda, disponibilização e a realização do acesso seguro, rápido e eficiente dos dados para a partilha em acesso aberto, sejam realizadas de forma essencial. Essa infraestrutura necessita estar interligada ao ambiente tecnológico e computacional para a gestão e preservação dos dados e de outros materiais importantes para a investigação. Nessa direção, as infraestruturas, que são aqui representadas pelas ferramentas e serviços, precisam que a governança da universidade utilize investimentos financeiros, políticos e de capacitação para as tornar articuladas e seus elementos interoperáveis e sustentáveis, mas também para minorar as tensões e conflitos que se apresentam no caminho para a realização das práticas essenciais para a RDM, que resultam em uma partilha adequada dos dados.

6.3.5. POLÍTICAS, NORMAS E ORIENTAÇÕES ADOTADAS

Perguntou-se aos investigadores sobre o conhecimento ou utilização de políticas, recomendações, orientações, acordos, padrões ou referências documentais institucionalizadas e específicas para gerir os dados de investigação, ou seja, qualquer tipo de documento que direcione os dados para serem geridos ou partilhados sob os cuidados necessários. As referências contidas nas respostas que se seguem, percorrem diversos tipos de documentos, conforme representado na figura abaixo:

Figura 39

Documentos citados em referência às políticas de RDM utilizadas



(...) as diretivas de operacionalização de todos esses dados e de gestão (...), na realidade como diretivas são inexistentes em Portugal (...) como as pessoas se veem órfãs de diretivas oficiais fazem o seu melhor (...) e vê como os outros fazem, (...) e implementam de acordo com as regras dos outros países (E1). (...) Há [no projeto] um documento com imensas páginas a dizer quando acontece isso o que é que se faz, quando acontece aquilo o que é que se escreve, para haver uma uniformização e sempre que surge uma situação nova, essa situação nova é adicionada (...), assim nós conseguimos uniformizar (E1).

(...) há (...) um código de conduta [que] abarca um conjunto alargado de temáticas que são importantes (...). Em termos de (...) documento de referência [nos] adaptamos e estamos mais robustos (E2).

Um conjunto de orientações (...), temos várias dimensões que se tocam e a parte da gestão de dados, as questões da ética e da integridade da investigação, também tem sofrido uma crescente regulamentação (...), a partir do momento que elas passam a ser regulamentadas, há um esforço de explicitação dessas práticas também (...), começamos por uma abordagem mais abrangente das questões éticas dos projetos que inclui, recolha e gestão de dados (...), proteção de dados (...)(E3). [A unidade de investigação] faz parte de um consórcio português (...) que é o arquivo das ciências sociais (...), está em ligação direta com o arquivo europeu e que está a definir (...) procedimentos próprios para a gestão de dados, com vista a preparação de dados que possam ficar depois em acesso aberto (...), no âmbito deste consórcio, também está a ser trabalhado um pequeno guia de trabalho para apoiar os investigadores no processo de preparação de dados (...) para que eles venham a ser disponibilizados, desde a fase da conceção do projeto (...). Em outra vertente (...), já preparamos um documento que ainda não é final (...), que dá conta de alguns procedimentos de segurança que nós assumimos, de proteção de dados pessoais que são recolhidos (...) durante a execução dos projetos e depois dos projetos terem terminados (...). Os documentos que estão a ser trabalhados de facto remetem (...) não só as práticas que (...) as equipas vinham adotando, mas também (...) documentos e regulamentação dos financiadores (...), ou seja (...), dar respostas aquilo que (...) é exigido e refletir aquilo que [a unidade de investigação] vê como boas práticas de investigação (E3).

Que eu tenha conhecimento não (...), aquilo que fizemos foi (...) os procedimentos normais de consentimento informado e de garantia que os dados vão ser anonimizados (E5).

Isso é deixado ao cuidado dos investigadores, até porque temos equipas muito heterogêneas, que trabalham com dados muito diferentes (...), é assumido que os investigadores tenham maturidade suficiente para saber lidar com os dados (...). Há colegas meus que trabalham com dados de saúde (...), antes deles o receberem passa pela comissão de ética do hospital (...), tipicamente os investigadores têm noção da importância dos dados que vão lidar e dos cuidados que devem ter com os dados (E6). (...) Normalmente há acordos de confidencialidade, os chamados *Non-Disclosure Agreements* (...), em que as pessoas que vão lidar com determinado tipo de informação (...) se comprometem a não divulgar (E6).

Ficou sempre nos protocolos que fizemos, assegurada a possibilidade de (...) usufruir e abrir os dados à comunidade (...). Sempre (...) tentamos negociar nos vários protocolos que fizemos (...), para a comunidade e produtos científicos a partir desses dados (...). Os dados eram de todos (...), mas sempre numa ótica de (...) transferência desse conhecimento para a comunidade científica, para a comunidade em geral (E7).

[Na] atividade de *crowdsourcing* (...) onde a gente dizia (...) precisamos da sua ajuda, se você consentir participar clique ok, a mesma coisa do termo de consentimento (...). As pessoas têm que estarem dispostas (E8).

Tem que haver sempre duas coisas que são essenciais, por um lado o projeto tem que ser submetido a comissão de ética do hospital (...), detalhados muito bem que dados vão ser adquiridos e em que condições (...), isto tem que estar perfeitamente claro (...), se os dados são apenas para esse estudo ou se os dados podem vir a ser acedidos para estudos posteriores. No passado isso não era uma preocupação (...) e agora há dados que não podemos usar (...) porque no passado não está lá escrito que esses dados podem ser usados no futuro (E9).

Quando eles se registam (...) há uma espécie de código de conduta (...), nós informamos aos utilizadores o que é que acontece com o conteúdo que eles nos dão (...), estabelecemos (...) regras gerais, relativamente aquilo que eles não podem fazer na plataforma e depois (...), quem criar a conta (...) concorda que (...) o conteúdo gerado pelos utilizadores, pode ser utilizado por outros utilizadores (...), o conteúdo que eles vão gerar é partilhado e fica como regra que lá é partilhado, a plataforma conserva a autoria e atribui a autoria, pois todo conteúdo recebe uma licença que é (...) CC BY NC 4.0 (...). E na política de privacidade definimos (...) [que] os dados de registo podem colher informações referentes ao seu computador, ao endereço de protocolo, ao tipo de navegador (...), informação que é própria do *google analytics* (...), fala-se dos *cookies* (...). O que garantimos é que toda a informação que é recolhida é exclusivamente para a utilização no âmbito das funcionalidades do site (...), uma das coisas (...) que está definida no nosso código de conduta é recolher ou armazenar dados pessoais sobre outros utilizadores, quer dizer que os próprios utilizadores assumem esse compromisso (...) de não recolher informação pessoal sobre outros utilizadores (...), essencialmente é aquilo que eles publicam em seu nome (E10).

(...) no *GitHub*, tem que declarar o tipo de licença de utilização dos dados, porque há um conjunto de licenças *standard* que são pré-negociadas. A licença que escolhi foi a MIT, a licença dos dados é diferente (E11).

As políticas que orientam as práticas para a RDM e a partilha dos dados são amplamente referenciadas na literatura (Asher et al., 2013; Borgman, 2018; Elliott & Resnik, 2019; Shearer, 2015). No entanto, este ator parece ter uma abrangência de aplicações quase ilimitada, pois podem ser estabelecidas em documentos para tratar da gestão dos dados pela própria UI&D (E1), mas também em códigos de conduta (E2 e E10),

orientações (E3), estabelecimento de termos (E5 e E8), códigos éticos (E6 e E9), protocolos (E7) e licenças para utilização dos dados (E10 e E11). No caso destes exemplos extraídos das entrevistas, pela característica dos entrevistados, são políticas mais direcionadas ao âmbito da investigação com ênfase na proteção dos dados, com o objetivo de adequação à regulamentação dos financiadores, da investigação e para garantir o acesso aberto.

No entanto, pode-se utilizar e conceber documentos orientadores para quase todas as atividades da RDM, de políticas para a partilha e preservação, até infraestruturas relacionadas ao acesso aberto, oriundas das agências financiadoras, órgãos reguladores ou construídas ao nível de governo dos países, mas também provenientes das universidades, das UI&D ou mesmo na própria investigação. Além das políticas formais, as práticas e regras informais são realizadas, isto porque segundo Borgman e Bourne (2022) as comunidades individuais têm seus próprios padrões e práticas variadas para o que é “comumente aceito”, tendo em vista que essas práticas evoluem ao longo do tempo, variam de acordo com o tipo de dados e podem não ser formalmente documentadas. Obviamente que todas essas possibilidades trazem consequências, por isso os autores chamam atenção para uma tendência destas elaborações serem “vagas, aspiracionais e flexíveis”, tão eficazes quanto sua implementação e conformidade subsequente. Por outro lado, quando planejadas e organizadas de modo pertinente, as políticas podem funcionar como um agregador de princípios e práticas que auxiliam os investigadores a lidar com os dados gerados no ambiente e as peculiaridades de que permeiam as suas práticas para a gestão e a partilha.

Portanto, a falta ou desconhecimento de uma política em nível da universidade que possa organizar e funcionar como orientação para a RDM, pode gerar regras e formulações de políticas locais ou das próprias UI&D, vistos e comentados nas últimas transcrições e reforçado no extrato abaixo:

Estamos a definir o que são os princípios orientadores base, que não podem deixar de serem cumpridos, mas deixando alguma margem de flexibilidade mesmo nesses modelos para abarcar essa complexidade disciplinar e metodológica (...), por exemplo (...), com recolha da imagem ou áudio temos obrigatoriamente duas opções (...), registo (...) para posterior análise e temos projetos em que esta recolha resultará ela própria em produtos da investigação do projeto, seja documentário (...), seja (...) projetos que utilizam (...) história oral (...), a recolher testemunhos que ficam depois em arquivos (...) e ficam disponíveis (...), aqui não há anonimato, aqui não há destruição dos dados recolhidos (...), o que estamos a tentar fazer é que os nossos documentos orientadores acomodem essa diversidade e não limitem (E3).

(...)o documento que estamos a trabalhar pretende uniformizar as regras de armazenamento para garantir que (...) esses dados brutos não se percam de repente e que possam ser mantidos até o momento em que supostamente possam ser destruídos (...), é um caminho mais longo que teremos que percorrer (E3).

Nós vimos exemplos [de políticas] (...), nos casos dos códigos de conduta eu vi alguns exemplos, no caso de política de privacidade meu colega da informática também viu alguns exemplos (...), os modelos que temos aqui estão em boas práticas de formatos internacionais (E10).

As regras formuladas nas investigações são necessárias e válidas quando se percebe que o ambiente da RDM, assim como a investigação, passa por transformações sob o escopo da ciência aberta. Porém, é necessário que estas transformações sejam direcionadas por políticas amplas, quer para a partilha dos dados em acesso aberto ou para outros aspectos no âmbito da RDM. A amplitude das políticas é pertinente porque não se pode acomodar toda a diversidade de práticas em documentos isolados. Assim as concepções e usos devem ser organizados nos níveis global e local, de modo que sejam complementares e não sobrepostos. Tais níveis, usos e concepções, mesmo difíceis de avaliar, são cruciais, tendo em vista os aspectos de tensões que podem surgir no ambiente (Borgman & Bourne, 2022).

É certo que as políticas podem direcionar as investigações nas universidades, com destaque para os consórcios e normas mandatórias das agências financiadoras, que repercutem na investigação e consequentemente na RDM, como a que entrará em vigor em 2023 do NIH (*National Institutes of Health*) dos Estados Unidos, que determina que todos os pedidos de subsídios para projetos que coletam dados de investigação, devem incluir um plano de gestão e partilha de dados, que contenha detalhes sobre os *softwares* ou ferramentas necessárias para analisar tais dados, quando e onde os dados brutos serão publicados e quaisquer considerações especiais para os acessar ou os distribuir (Kozlov, 2022). A força deste tipo de política pode ser percebida em relatos dos investigadores, como os que seguem.

Um consortium (...) que (...) anda a definir os padrões (...) como a introdução dos dados, as linhas mestras a que deve obedecer às bases de dados, todos os processos e intercambio de informação, de tal maneira que a partir de determinada altura (...) quem (...) não seguia esse modelo (...) mais tarde ou mais cedo deixavam de ter interoperabilidade (E1).

Nos últimos dois a três anos isso começou a ser mais frequente. Nós trabalhamos com a agência europeia de direitos fundamentais (...) e tivemos que preencher todo um inquérito onde explicamos bem quais os procedimentos que a gente garante a proteção da informação e salvaguarda dos acessos restritos ao armazenamento da informação (...) e depois da utilização da informação, ao fim de dois, três, quatro anos após os resultados finais (...), então procede-se a sua eliminação (...). Em Portugal isso foi mais relevante a partir do momento em que muitas instituições (...) começaram a receber financiamentos internacionais (...) onde houve de facto uma maior pressão por

procedimentos (...) que salvaguardem todas as questões éticas que possam estar envolvidos nos projetos (E2).

E há (...) iniciativas da comissão europeia e dos Estados Unidos para que os dados públicos, sejam disponibilizados de forma aberta (E6).

A (...) exigência é que alguns *outputs*, e um dos *outputs* do projeto era a base de dados, ficasse acessível ao público (E7).

Esse financiamento por regras da união europeia obriga que tudo que é produzido seja disponibilizado em acesso público (E10).

O entendimento da importância das políticas mandatórias é ratificada no resultado da análise da literatura, onde é definido que o processo para uma partilha adequada dos dados deve abarcar as orientações e as exigências das políticas. De forma geral, nas concepções resultantes desta análise, as políticas são destacadas como um dos atores a serem necessariamente articulados, mas também em termos de proteção legal orientado pelas práticas essenciais de gestão. Assim, as políticas, regras e documentos existentes ou a serem elaborados que envolvem a RDM, devem ser prioridade para as práticas que possibilitam a gestão e a partilha dos dados na universidade, não somente pela questão dos financiamentos, mas porque esse ator juntamente com outros que devem ser articulados, influenciam a RDM de maneira determinante para o sucesso, ou como barreira para a partilha dos dados, tendo como consequência os problemas que o modelo daqui resultante se dispõe a minimizar.

Inquiridos especificamente sobre o uso de um plano de gestão dos dados, a maioria das respostas convergem para o conhecimento sobre o que seja o documento, mas sem a sua utilização de modo formal. Quatro entrevistados falam sobre o assunto do seguinte modo: um informa que não utiliza na investigação, embora tenha essa prática informal, os outros três indicam e comentam a possibilidade de utilizá-lo, conforme a transcrição dos relatos.

Não existe um plano de gestão de dados (...), não existe nada formal ainda, ele existe de uma forma oficiosa, nós sabemos que temos que conservar os dados, fazemos nossas cópias da forma que entendemos que devemos fazer (...), com base na nossa experiência pessoal, com aquilo que vemos os outros a fazer, mas não temos um plano escrito formal para a gestão dos dados (...). Eu sei que a entidade *Research Data Alliance*, tem aqui um nó e que eles promovem essas formações, mas ainda não houve oportunidade de ir a nenhuma e não temos isso formalizado (E1).

Cada projeto terá um plano de gestão de dados (...) porque todos os projetos são diferentes, todos os projetos vão tratar assuntos diferentes, vão lidar com recolha e tratamento de dados de forma diferentes (E3).

Eu sei que hoje em dia nas candidaturas aos projetos é obrigatório haver um plano (...), mas esse momento muitas vezes é descurado (...). Nós já tivemos vários consultores no projeto, (...) numa das últimas reuniões (...), um professor americano que trabalha (...) nas humanidades digitais ele dirigiu justamente a conversa para esse tópico (...), a questão de um plano de gestão dos dados, de

preservação a médio e a longo prazo e nós temos isso como um projeto mais ou menos definido (...). Tenho até um guião de duas páginas que falo (...) que esse plano de gestão de dados fique garantido a médio e a longo prazo (E10).

Há duas coisas que são muito importantes a fazer, uma é que é preciso haver uma análise formal do universo que se vai tratar e explicar (...), que tipos de documentos são, que características têm (...), [outra é] uma análise a longo prazo em termos técnicos de um plano de gestão (E11).

O plano de gestão de dados foi citado no decorrer deste trabalho como um documento necessário para a RDM na partilha em acesso aberto, por exemplo, na pesquisa bibliográfica grande parte dos artigos recuperados discorreram sobre o assunto, bem como nos modelos estudados (DataONE, 2021; Jisc, 2021; University of Virginia Lybrary, 2022). O plano de gestão de dados é um documento necessário porque, além de ser parte da exigência de financiadores (Kozlov, 2022), descreve como serão tratados os dados durante um projeto e o que acontece com os dados após o término do projeto, normalmente abarcam parte, ou todo o ciclo de vida dos dados, assim como as práticas para preservação, infraestruturas tecnológica, partilha e outras inúmeras possibilidades (Michener, 2015), isso a depender do tipo de plano a ser utilizado. Normalmente estes planos expressam as orientações previamente definidas pelas políticas institucionais e modelos dos financiadores, mas sobretudo devem refletir o tipo de investigação e suas especificidades.

A intencionalidade da utilização de um plano de gestão de dados demonstrada nas respostas e o não uso formal pelos investigadores indagados, pode indicar que sob a tendência global de investimentos para que haja a partilha dos dados em acesso aberto e também um plano de gestão de dados, as UI&D, as investigações e os investigadores devem ser preparados através de políticas promocionais, infraestruturas e práticas para que a elaboração deste plano para a RDM seja recebido de modo a gerar o mínimo de barreiras. Isto porque o estabelecimento de mais atribuições e responsabilidades, como a elaboração de um plano de gestão de dados devem vir acompanhados de esclarecimentos e suporte para que a sua repercussão seja aceitável, do contrário poderá ser tratado como mais um documento a ser preenchido para obtenção de recursos financeiros, o que provavelmente pode repercutir de forma direta e negativa na partilha e na reutilização dos dados.

Outro item que envolve as regras e as normas para a RDM são as questões legais, como os direitos autorais e a propriedade intelectual, também foi abordado com os entrevistados. Ao serem perguntados se existe no âmbito da investigação, algum

documento, norma ou orientação técnica relativa a tais questões, os investigadores respondentes discorreram sobre o tema nas seguintes considerações:

Não na generalidade (...), o que acontece é que [a unidade de investigação] tem uma forte componente sociojurídica e também temos na nossa equipa alguns juristas, (...) que nos ajudam a clarificar (...). Em outros projetos com parceiros que tem uma dimensão tecnológica mais forte e quando há questões relativas à propriedade intelectual (...) ou registo de patente, ou gestão da inovação (...) em que não trabalhamos tanto (...), nesses casos (...), socorremos dos outros parceiros do consórcio que têm essas competências já mais refinadas (E3).

(...) quando estamos a falar de dados que envolvem utilizadores (...) já existe regulamentação suficiente (...) com a regulamento geral de proteção de dados, que garante um conjunto de direitos aos utilizadores cujos dados são recolhidos (E6).

Nos protocolos com as câmaras e com as entidades externas uma das coisas que salvaguardamos sempre é o caráter não comercial dos dados (E7).

A questão legal exerce influência sobre a tomada de decisão em relação à partilha da produção científica, mas também sobre repositórios institucionais e na RDM (BOAI, 2002; Cox & Pinfield, 2013). Assim, conforme a legislação de cada país e os acordos internacionais, com destaque em áreas como direitos de propriedade intelectual e proteção da privacidade, este item da RDM afeta diretamente as práticas para a partilha, de forma que devem ser considerados nos acordos de acesso aos dados (OCDE, 2015). É possível perceber que existe uma falta de entendimento e de apoio específico para a questão, uma carência verificada não somente nestes excertos, mas que também pode ser vista na seção deste capítulo sobre colaboração, uma vez que quando indagados sobre quais outros atores poderiam ajudar na RDM, a área jurídica foi apontada como necessária.

Possivelmente a necessidade de apoio e a falta de informação sobre a abrangência legal que a RDM e a partilha dos dados em acesso aberto requerem, refletem a falta de uma estrutura organizacional para esclarecimento desta questão, com o respectivo suporte jurídico, de forma a abarcar áreas gerais, mas também específicas da RDM. São questões que abrangem um arcabouço referencial de princípios e práticas oriundos da ciência aberta, referente às leis de privacidade e segurança, licenças, contratos de confidencialidade, normas, códigos de ética e de condutas e outros mecanismos semelhantes que direcionam às proteções legais necessárias para que o investigador possa partilhar em acesso aberto. Portanto, a partilha em acesso aberto pode proporcionar vulnerabilidades quando não observada a conformidade jurídica para os direitos existentes, concorrendo para a

insegurança que faz com que o investigador tenha dúvidas para partilhar em acesso aberto, juntamente com as consequências desta exposição dos dados.

A análise do resultado da literatura corrobora neste contexto através das práticas essenciais, com a indicação da construção de uma proteção legal sob a RDM, que deve ser direcionada para dirimir as vulnerabilidades e proporcionar segurança para o contexto da investigação e dos seus atores. São muitos os itens a serem observados que viabilizam a proteção da propriedade intelectual, tais como: os direitos dos autores e investigadores às suas produções científicas e de serem reconhecidos e citados, mas também estão inclusos o tratamento e os limites para acesso ao material da investigação, como os dados sensíveis, sigilosos e de informações pessoais, de cunho privados ou públicos que possam incorrer em danos particulares, econômicos e sociais. Nesse sentido, as questões legais é um dos itens necessários para a RDM sob a partilha dos dados que deve ser explorada.

Os investigadores também puderam responder sobre a utilização da documentação que permite descrever ou contextualizar os dados, especificamente em relação aos metadados. A maioria dos respondentes não expuseram o assunto por não ser uma prática realizada na investigação, enquanto outros assinalaram questões como as que segue abaixo.

(...) implica a codificação da informação, colocar um descritivo de que informação é o que é, onde que foi retirada (...). Se multiplicar isso por entrevistas, observações, registos fotográficos, vídeos, estamos a fazer um efeito multiplicador da informação que não é disponibilizado (E2).

(...) isso é uma boa questão que se levantou quando eu ouvi falar desta possibilidade dos dados serem reutilizados (...), já conversamos sobre a ideia de adotar um tipo de formato (...) que seja mais persistente (...), mas nunca avançamos muitos nesse sentido (...), no futuro será um caminho (...) [que] tem que ser trilhado (E7).

Sim, tem que haver sempre, senão ninguém sabe o que fazer com os dados (...), a equipa de investigação disponibiliza os dados, tem sempre que acompanhar uma pequena descrição, mas não é nada muito complicado (...), é dizer que variáveis (...) em que condições (...), mas sim (...) há sempre um documento explicativo dos dados (E9).

(...) na possibilidade de partilharmos os dados de uma maneira compreensível, partilharmos a interpretação dos dados (...), mas partilhar de uma forma que outra pessoa possa (...) ver isto de outra maneira (E11).

A concepção e uso dos metadados, remetem a necessidade de disponibilizar a informação pertinente para identificar os dados, a qual tem por objetivo preservar e deixá-los acessíveis para ser processado e utilizado com significado no futuro (Hachen et al., 1994). Tal importância traz a obrigatoriedade do uso dos metadados para a RDM de modo

a tornar os dados interpretáveis, reproduzíveis e reutilizáveis. Mas também amplia o escopo do trabalho, pois para que os dados sejam reutilizáveis nesse contexto, é necessário fornecer uma variada e extensa documentação de protocolos, modelos, ferramentas e *softwares* (Borgman & Bourne, 2022). Isso significa que os metadados podem variar de acordo com o objeto a ser representado e com a especificidade da investigação, contudo é necessário que sejam legíveis por sistemas e verificados padrões para a interoperabilidade na infraestrutura tecnológica, o que demanda uma maior especialização.

O trabalho a ser realizado na concepção de metadados relevantes, ganha na figura da biblioteca um ator importante, pois de acordo com Soehner, Steeves e Ward (2010) os metadados constituem uma área da *e-Science* onde as bibliotecas encontram uma afinidade natural. Borges (2006), afirma que o papel da biblioteca será o de servir de mediadora, um papel particularmente relevante para facilitar o processo da recuperação da informação. Se também forem aqui recuperadas as transcrições das respostas da parte sobre a colaboração, a biblioteca é vista como um dos atores que pode colaborar em termos de padrões e metadados (Tabela 22), mas não só por isso, neste contexto, as bibliotecas também aparecem na literatura como órgãos centrais (Cooke, 2007; Jones, Pryor & Whyte, 2013).

Além de presente como um possível serviço das bibliotecas, os metadados estão em diversos ambientes e práticas que envolvem a RDM, ou seja, são direcionados pelas políticas (OCDE, 2021), representados nos modelos de ciclos de vida (Higgins, 2008), bem como constante nos princípios FAIR, este último muito bem ancorado nos metadados (Wilkinson et al., 2016). Assim, no entendimento de que os dados de investigação não têm valor sem seus metadados e documentação apropriada, para descrição do contexto e das ferramentas usadas para criá-los, armazená-los, adaptá-los e analisá-los (Kindling, 2013), são várias as práticas para a RDM que contam com os metadados como elementos obrigatórios para assegurar que possa ser realizada uma partilha adequada dos dados, como, por exemplo, a interoperabilidade, a preservação e a reutilização dos dados. No entanto, para que isso aconteça é necessário dispor das condições técnicas e infraestruturais, de modo que o investigador possa ser apoiado por meios articulados e que os metadados sejam criados a contento das especificidades da investigação, resultando na partilha e reutilização dos dados, a partir do ambiente da RDM na universidade.

6.4. PARTILHA DOS DADOS DE INVESTIGAÇÃO EM ACESSO ABERTO

O terceiro e último bloco temático, obtém informações sobre os hábitos dos investigadores em relação às práticas para a partilha dos dados de investigação em acesso aberto, procurando também as percepções no que diz respeito aos fatores que habilitam ou restringem esse tipo de partilha no âmbito da investigação. Completa-se também com a realização deste bloco temático, o objetivo específico seis da tese.

6.4.1. HÁBITOS E PERCEPÇÕES PARA A PARTILHA DOS DADOS EM ACESSO ABERTO

Os investigadores foram perguntados se disponibilizam os dados de investigação em acesso aberto. A maioria dos entrevistados ainda não realizam este tipo de partilha, as respostas que foram positivas se deram no sentido de como acontece tal disponibilização, os recursos utilizados e as justificativas para a partilha, como segue na forma abaixo.

Nossos dados são de acesso livre (...), estamos em uma plataforma global (...), fizemos a disponibilização dos dados nessa plataforma (...), estão lá mais de noventa mil registos (...), mas (...) o formato de dados que eles usam não incluem toda a informação (...), há informação que não migram para lá porque não é informação que interessa a eles (...). A migração é feita (...) no modelo (...) é (...) mais usado para migrar dados da biodiversidade (...), o acesso é livre e sempre será (...), é um consórcio em que a adesão é feito a nível dos governos dos países (...). Há, portanto, uma certa garantia que os dados serão abertos e acessíveis a todo mundo (...), é provavelmente a base de dados mais FAIR que existe (E1).

(...) disponibilizamos em uma plataforma para ser de acesso público e gratuito, para que outros investigadores possam utilizar no futuro (...), estou (...) também [em] um outro projeto de investigação financiado pela FCT (...) [que] após terminarmos (...) é minha intenção que a base de dados vá para a PASSDA, porque é uma área temática que eu posso explorar no futuro (E2).

[os dados] são postos em bases (...) acessíveis ao público (...), entram depois em algumas bases internacionais, são partilhados (E4).

(...) se calhar até existe um sítio qualquer onde esta informação está, mas a minha própria gestão de dados já não me permite ir lá ou já não encontro, já não sei onde é que está (E5).

(...) uma das nossas preocupações foi de torná-las abertas ao público, porque (...) temos (...) pavor [das] bases de dados às quais não conseguimos ter acesso e são financiadas com dinheiro público (...), não foi o caso das nossas, mas achamos interessante fazer essa contribuição e elas estão (...) abertas ao público (...). Não têm restrições de uso ou de utilização dos dados, completamente livre, o que o utilizador não autorizado não faz é introduzir novos dados (...). O código também é código aberto (...), tem tudo depositado no *GitHub* (E7).

Mas a ideia toda é que esteja aberto, disponível e que seja pesquisável por pessoas que não seja especialistas em linguística de *corpus* (E8). Foi colocado no *GitHub*, o pessoal das ciências informáticas (...) fica com a parte do *GitHub* e a parte do *game design* vai também ficar disponibilizado (E8).

O [projeto] ficou em acesso aberto em 2005 (...). A biblioteca nacional (...) esteve a digitalizar todos os papéis (...) e eu também consegui que a biblioteca nacional nos autorizasse a disponibilizar (...), temos esses materiais, temos a investigação que nós produzimos e toda a propriedade intelectual que foi criada por nós, depois temos também o código em acesso aberto no *GitHub* (E10).

(...) o mais importante da nossa conversa é conseguir (...) fazer isso de uma maneira que possa ser reproduzida por qualquer pessoa. Portanto, qualquer pessoa pode abrir o VS Code, ir ao repositório que está no GitHub, fazer *download* do ficheiro Excel original, fazer o processamento que extrai estas informações e depois tem uma série de *notebooks* organizados (E11).

A partilha dos dados em acesso aberto por esses investigadores acontece em plataformas e ferramentas disponíveis por instituições de áreas temáticas ou parceiras, bem como a partilha do material no *GitHub*, em acesso aberto na *Web*. As justificativas apresentadas são para que os dados estejam em acesso livre e gratuito ao público; dentro dos princípios FAIR; possam ser explorados no futuro; sejam pesquisáveis também por não especialistas e porque as investigações são financiadas com dinheiro público. De forma geral, as respostas concordam com Piwowar et al. (2008), e Sayão e Sales (2020), no sentido de que as opções para partilhar os dados de investigação estão em locais como centros de dados, repositórios institucionais, temáticos, multidisciplinares ou desenvolvidos por projetos específicos. Nesse sentido, os dados partilhados, que são os processados e não os dados brutos da investigação, estão em repositórios de dados para acesso e reutilização, possibilitando que parte das justificativas dos respondentes para a partilha possam ser realizadas.

Neste contexto da partilha dos dados em acesso aberto, cabem duas observações quando no confronto com o resultado da análise da literatura. A primeira vai no sentido de que a ação da partilha, que nesta investigação foi desenvolvida sob a ideia do cuidado que se deve ter antes deste ato, requer uma série de práticas para garantir que a partilha não se torne um fim em si mesmo (Paschetto, Randles & Borgman, 2017). A garantia de que os dados podem ser reutilizáveis, além de ser um princípio central da ciência aberta e objetivo da RDM (Palmer, Weber & Cragin, 2011; Cox & Pinfield, 2013) indica a realização de uma partilha adequada, pois o trabalho que envolve a preparação para que os dados sejam reutilizáveis passa pela observação de vários princípios e práticas da ciência aberta, essenciais e necessárias para obter as vantagens dos dados abertos e seu uso na ciência e na sociedade. Assim, a partilha não deve ser analisada de forma imediata no seu ato, mas deve ser vista em conjunto com o contexto e o ambiente da investigação e do investigador.

A segunda observação é a organização para que a partilha seja direcionada à reutilização (Borgman, 2007; Niu, 2009; Piggot, Hobbs & Gammack, 2001). Existe uma série de dificuldades a ser considerada para a reutilização dos dados, no contexto da

universidade, pois, como visto nas transcrições, a partilha pode ser realizada em locais muito diferentes entre si. Saber quais são esses locais, repositórios ou serviços e onde os dados estão partilhados, bem como garantir que a partilha seja adequada, faz parte do escopo de utilidade que uma RDM na universidade pode proporcionar, indicando os caminhos para a isto. Portanto, no entendimento de que já existe uma RDM em cada investigação, onde parte das investigações partilham seus dados em acesso aberto, a universidade tem como necessidade e desafio também saber onde estão os dados já partilhados e prover os meios necessários para que a RDM esteja sob uma orientação geral, porém sem perder as especificidades concernentes da temática e as diferenças disciplinares de cada investigação.

Em uma pergunta decorrente da anterior, foi questionado aos entrevistados os possíveis motivos existentes em termos de dificuldades, interesses e preocupações, segundo as percepções e a experiência dos investigadores respondentes, para que não se partilhassem os dados de investigação em acesso aberto. Mas também qual a compreensão em relação ao seu ambiente para que tal partilha não aconteça. Parte dos motivos estão nomeados na figura 40, o contexto destes motivos e outras especificações seguem citados após a figura.

Figura 40

Motivos e percepções para a não partilha dos dados em acesso aberto



(...) tem a ver com um fenômeno geral que é a democratização e a democratização da utilização (...). À medida que nós facilitamos o acesso e deixamos que qualquer pessoa interfira com (...) os dados, nós corremos cada vez mais riscos (...), as coisas complicam-se (...) e também traz dissabores com a utilização dos dados (...) que estão *online* (...), agora começa a colocar-se um problema (...), é que durante muito tempo as pessoas que chegavam a esta informação, tinham que saber como chegar lá, isto não é fácil (...), de maneira que nunca houve preocupação (...), mas hoje com as redes de informação que nós temos (...) quem vai à Internet (...) nem sempre tem a preocupação de informar que utilizou os exemplares (E1).

(...) obrigou os investigadores a ter uma maior carga burocrática, de ter os passos todos muito bem-organizados e ter (...) cuidado na forma como armazenar a informação, como tratar a informação (...), mas havia alguns argumentos, alguns motivos (...) que eram invocados para essa restrição da disseminação da informação, nomeadamente para poderem ter o poder exclusivo para outras publicações futuras, para manter seu nicho (...) de uma forma mais controlada (...), para menor controlo da investigação, ou seja (...) é mais difícil controlar nos processos de verificação da qualidade (E2). (...) também há uma dimensão cultural que é a parte dos investigadores ainda não estarem muito disponíveis para (...) partilhar a informação que recolhem no âmbito dos seus projetos de investigação, como ainda não beneficiaram muito provavelmente dessa partilha, ainda não estão muito disponíveis para partilhar (E2). (...) isso é uma questão geracional, não só (...), mas (...) nos anos setenta, sessenta até oitenta (...) no sentido da comunidade ser mais pequena, estarem mais isolados (...) havia um maior fechamento em termos dos processos de investigação e do controlo da informação que se produzia (E2). (...) Um investigador que ao final de seu projeto queira disponibilizar as (...) entrevistas que fez (...), necessita garantir que estas entrevistas estejam anonimizadas (...) [com] transcrições bem efetuadas (...) não [existe] (...) essa capacidade, pessoal, profissional ou financeira para disponibilizar essa informação sem que antes toda ela seja trabalhada (...) [na unidade de investigação] que tem oitenta a noventa projetos financiados, em curso, significa que estamos a falar de uma escala de informação muito grande (...) cujos materiais da investigação (...), a maioria não é disponibilizada publicamente por esta dificuldade (...), é muito complicado que estes materiais sejam disseminados e reutilizados em outros contextos (E2).

(...) há ainda uma certa resistência da parte de investigadores que trabalham com esses tópicos (...). Uma grande parte dos investigadores que trabalham com metodologias qualitativas e temas deste tipo entendem que disponibilizar os dados, mesmo que trabalhados e tratados, é uma espécie de traição (...) [da] confiança estabelecida com os participantes (...), acho que há um caminho longo a percorrer para que os investigadores e investigadoras que trabalham esses temas consigam (...) ver a relevância e também a ética de disponibilizar os dados (E3). (...) Porque o tratamento da informação muitas vezes não é suficiente para anonimizar por completo (...), porque realmente não tem tempo, não têm os recursos humanos suficientes para conseguir fazer um trabalho (...) de edição, para que eles fiquem (...) disponíveis (E3).

A informação é conhecimento e poder, logo quem detém os dados (...). Há centros que eventualmente não querem dar alguns dados a outros centros, até para a própria universidade, e da posse deles pensa-se erradamente (E4). Sendo objetivo (...) não sei se os outros colegas que não trabalham conosco têm a perceção da importância da área técnico-científica que trabalhamos(...). Alguns colegas sim, mas outros eventualmente não (...), não têm a perceção da importância (...), para mim (...) os cientistas das várias áreas não falam, não trabalham e não pensam em conjunto (...), eu penso que os cientistas estão fechados (E4).

Corre sempre o risco em ser disseminado onde nós não queremos (...), a partilha e a disponibilização pode levar a um sítio em que nós não queremos, isso é uma coisa que me preocupa, se os dados vão cair em mãos alheias que não os tratam como deve ser. E em última instância, quem está na origem desses dados deveria poder dizer (...) não quero (...), anonimizados ou não, os meus não (E5).

Eu diria que depende dos dados, muitas vezes esses dados de colaboração são dados (...) das empresas (...). Em projetos de *smart cities* pode ter repositório de dados abertos, por exemplo, com dados de empresas de telecomunicações que ajudam a perceber a mobilidade na cidade (...), mas o

que é que a empresa tem a ganhar com isso? (E6). (...) Há sempre uma resistência em publicar esses dados (...), a menos que sejam da administração pública (E6).

Acho que o foco está em (...) cumprir as metas definidas e não tanto nos dados em si (...). No decurso do projeto (...) as pessoas estão demasiadas envolvidas no processo para poderem refletir (...). Pelo menos no âmbito dos historiadores (...), [a comunidade] não está (...) sensibilizada e desconhece todos esses debates (...), não é uma preocupação que esteja no horizonte das pessoas (...), não só na questão dos dados (...), mas na partilha da informação (...). Na questão dos dados de investigação a literacia é muito baixa mesmo (E7). Estamos a falar de uma tribo muito individualista, com uma noção de que o trabalho é meu e eu não partilho com meu colega do lado (...), os dados são meus, eu fico com eles, aquilo que eu trabalhei é meu, eu vou rentabilizar isso para minha produção científica, mas não partilho (...), se alguém quer trabalhar a mesma coisa vai ter que começar do zero apesar de eu ter ido até ao vinte (...) e durante anos isto aconteceu até com projetos financiados, há imensas bases de dados que foram produzidas no âmbito de projetos da área da história que ninguém nunca teve acesso, nunca ninguém viu, a não ser os investigadores desses projetos (E7).

Eu não sei se na altura a pessoa está mais preocupada em ter os dados para seu próprio projeto. Tentar fazer com que os dados sejam públicos é muito raro que isso aconteça, quando muito tentar que os dados possam ser acessíveis para estudos posteriores ou para colegas que vão analisar os dados segundo outras abordagens (E9).

As coisas têm que coexistir, as universidades têm a obrigação de ter algum processo, assim como fizeram os repositórios que são coisas de inestimável valor (...), haver alguns sistemas de repositórios de dados é importante, o que eu acho que é perigoso é não reconhecer a diferença. (...). No repositório de dados, toda a gente vai lá pega os dados e faz alguma coisa com ele, isso é razoável? Acho que não é razoável, a não ser que se incorpore coisas (...), que mostram poder suportar cooperação em larga escala, complexa, entre comunidades de pessoas e que na minha opinião vão se tornar cada vez mais fáceis de se fazer (...). Não é só eu saber que os dados estão lá em um sítio qualquer e que estão categorizados (...), mas na prática, qual é o custo e o tempo necessário para alguém pegar nos dados e reanalisar com uma hipótese diferente (...), a velocidade e o custo, no sentido de maior velocidade e menor custo com que é possível reutilizar os dados (E11).

Ao observar a figura 40 e as transcrições acima, percebe-se que os fatores para não partilhar os dados são muitos e estão concatenados com quase todos os blocos anteriores, sejam políticas, infraestruturas, questões cognitivas, recursos, entre outros. Ou seja, as barreiras estão espalhadas no ambiente da RDM na universidade, prontas a resultar em uma partilha inadequada ou mesmo na não partilha dos dados, são pontos que em sua maioria têm pertinência e estão entre os problemas que esta investigação tem como parâmetro de trabalho. Os problemas se desdobram à medida que tenta se observar mais de perto os motivos dos entrevistados para a não partilha dos dados. Assim, atentar para cada aspecto identificado é um trabalho volumoso, mas necessário e requer o envolvimento de vários atores e uma continuidade de trabalho para minimizá-los.

É também interessante perceber que, embora a lista de respondentes não seja extensa, os motivos apresentados estão em boa parte consonantes com o estudo de Veiga (2017), que apresenta fatores preocupantes por parte dos investigadores que impedem a partilha em acesso aberto, bem como abrange as categorias de comportamentos: individual,

institucional e técnico (Tabela 10). Acrescidos ao ambiente diverso e heterogêneo da universidade e dos dados produzidos, parece que tais dificuldades podem se multiplicar a ponto de inviabilizar qualquer iniciativa ampla a nível da universidade. No entanto, Borgman e Bourne (2022), indicam como deve-se agir sobre tamanhas desigualdades no ecossistema dos dados, expressa na frase: “pense globalmente e aja localmente”. É nesse sentido que o resultado da análise da literatura aqui realizada é desenvolvida para a RDM que resulta em uma partilha adequada dos dados, com atores articulados e no uso de práticas essenciais. Pois as possíveis barreiras, bem como uma gama de especificidades que devem ser levadas em conta, sob uma padronização necessária, fazem com que as soluções devam ser pensadas de modo geral a partir de práticas específicas e locais.

Quando, contrariamente à pergunta anterior, os entrevistados foram questionados de modo específico sobre os motivos que justificam, ou a razão, para que se deva partilhar os dados de investigação, a percepção da maioria dos investigadores converge para o interesse na disponibilização dos dados em acesso aberto. Chama atenção que parte dos relatos parecem ter seus fundamentos em questões negativas para fundamentar as positivas. Tais percepções estão representados na figura abaixo, as transcrições seguem logo após.

Figura 41

Motivos e percepção para a partilha dos dados em acesso aberto



Um botânico, essencialmente é uma pessoa que quer partilhar (...), está nos nossos genes (...), não são como esses dados da farmacologia (...) ou algumas coisas da engenharia em que há um grande problema de confidencialidade (...), nada, ciência totalmente aberta (E1).

Não fico preso à ideia de que este é o meu domínio e, portanto, (...) vou partilhá-lo, por acaso se houver mais gente a trabalhar na área, mais rica é a informação que depois se trabalha de uma forma coletiva (...). O que nós sentimos é que há pouca gente a trabalhar nesta área (...) e por isso gostaríamos que houvesse mais gente, e que outros que trabalham em áreas próximas possam também utilizar a informação. (...) E se nos preocupamos com a transformação da sociedade em uma sociedade mais igualitária, mais coesa com mais qualidade de vida, então essa informação tem que ser partilhada (...). Hoje, até pela projetificação da ciência e eu diria (...) por uma tendência mais neoliberal (...) e portanto (...) sob uma pressão muito grande para publicar, produzir, mostrar resultados, indicadores que os tornem mais competitivos para os financiamentos da investigação e para conseguir um contrato de trabalho permanente (...), seja numa universidade, seja num centro de investigação (...), isso levou que as gerações mais jovens tenham uma outra abordagem para a investigação, que é investigar e publicar, a pressão pela publicação é tão grande que logicamente força essa partilha (...), e por essa via, a informação resultante da investigação é mais partilhada (E2). Tem tido uma grande mudança nos últimos anos para a necessidade da partilha e a disponibilidade em partilhar, assim como as políticas europeias e mesmo nacional de acesso aberto também potenciaram e consciencializaram a comunidade científica para que esse processo ocorra mais facilmente (...). Começa a ser já uma necessidade que se sente para a partilha de informação, resultantes de financiamentos públicos, portanto se é público acho que é um dever de partilhar, mas com os cuidados naturais da informação (...), é melhor partilhar para salvaguardar os indivíduos de plágios, por exemplo, do que propriamente estar a esconder a informação que acaba por ser sempre pública de alguma forma, quer seja pelas publicações, relatórios (...) que se fazem (...). Portanto, eu sempre fui mais favorável a partilha e a publicação (E2).

(...) É sempre considerada a possibilidade de partilhar os resultados do projeto (...), não é necessariamente partilhar os dados (...), é partilhar (...) materiais que tenham uma utilização prática depois (...). Tenho para mim que a ciência deve ser toda aberta (...) se os dados (...) forem quantitativos (...), acho que aqui não há grande problema (...), mas isto será projeto a projeto (E3).

É uma coisa que está cada vez na ordem do dia, em termos de ciência não é propriamente agora (...), [originada] principalmente nos anos oitenta com o aquecimento global, as alterações climáticas (...). Acho que isso é muito importante (...) a todos os níveis, entre os cientistas, entre os cientistas dos outros centros (...), o sair da comunidade e o devolver a sociedade (E4).

[Em relação à partilha] a perceção sobre o valor dos dados tem vindo a aumentar ao longo dos anos e quanto mais sofisticados são, por exemplo, os algoritmos que são utilizados para analisar os dados e descobrir padrões, (...), se nós tivermos muitos dados, à primeira vista podemos não ver nada, mas à medida que as técnicas computacionais são cada vez mais sofisticadas, o que não se vê à primeira vista revela padrões interessantes de consumo (E6).

(...) se (...) futuramente a universidade criar ou se houver uma infraestrutura supra universitária onde se possa estabelecer essa ligação e esse depósito (...), eu pessoalmente não vejo nenhum problema em depositar a base de dados que eu criei nessa infraestrutura (...), saber que tem que depositar (...) nos repositórios e que isso traz vantagens (...), acessibilidade é importante para a universidade, para a construção do currículo (E7).

Se a pessoa está a trabalhar num determinado assunto, o grande problema (...) é acesso a dados (...), precisamos dos dados (...) e ter acesso aos dados é o nosso grande problema (...), os dados existem, não estão públicos, estão aí nos hospitais, não estão disponibilizados. É evidente que eu vejo vantagens nos dados serem públicos, porque eu tentei resolver um problema com esses dados, se eu disponibilizar esses dados depois (...) podia acontecer que houvesse outros investigadores, com os mesmos dados, desenvolvessem soluções melhores ou complementassem aquilo que eu tivesse já desenvolvido (...), sinceramente não vejo problemas nenhum para que os dados não fossem disponibilizados em acesso aberto (...), o problema não é disponibilizar os dados, o problema é (...), no final para fazer um estudo clínico que custa duzentos, trezentos mil euros ou mais (...) e chega

ao final então quem quiser, usa os dados, isso é um problema (...), quem têm os dados não os disponibiliza porque lhes custaram muito a arranjar (E9).

E essa lógica da partilha eu acho que é muito importante, porque uma das coisas que eu acho que o meio digital faz é: tornar os nossos papéis mais dinâmicos, isto é, nós estamos menos restritos àquilo que a cultura tipográfica constituiu que é, de um lado está o autor, de outro está o editor, do outro está o leitor, o meio digital permite que os nossos papéis sejam muito mais fluídos, passamos de uma posição para outra e [a investigação] baseia-se um pouco nessa lógica (E10).

Partilhar os dados e tornar transparente o processo de decisão de inferência (...) é a revolução na possibilidade de partilharmos os dados (...), mas partilhar de uma forma que outra pessoa possa, sem um investimento extraordinário de tempo (...) ver isto de outra maneira (E11).

Na perspectiva de comparação com o resultado da análise da literatura, percebe-se que estas motivações expressas para a partilha são semelhantes a outras já identificadas (Veiga, 2017). Como já afirmado, embora os fatores negativos sejam importantes e consideráveis, verifica-se também nesses comentários que mesmo perguntados sobre os fatores positivos, recorre-se aos negativos como suas justificativas. Assim, ao comparar as causas negativas citadas anteriormente com estas motivações para partilhar em acesso aberto, juntamente com a literatura, pode-se entender que embora ressalte-se bastante os aspectos positivos e os investigadores reconheçam parte deles, de modo geral os aspectos negativos são mais enfatizados, principalmente nas entrevistas (Freitas & Leite, 2019). Um dos motivos pode ser porque ainda que direcionados pelos benefícios da partilha em acesso aberto, os fatores desfavoráveis estão originalmente na realidade da investigação e quando investigados sob este ambiente, são mais facilmente relatados e transpostos para a literatura. Nestas bases de percepção, os dois fatores estão interligados e em dependência, logo nem fatores negativos nem positivos podem ser vistos isoladamente, pois estão intimamente relacionados ao contexto da universidade e ao ambiente da investigação.

O contexto aparentemente com ênfase no negativo não significa que o investigador não deseja partilhar, porque, ao longo dos anos, a literatura tem mostrado consistentemente que a maioria dos investigadores é a favor e estão claramente interessados na partilha em acesso aberto (Dallmeier-Tiessen et al., 2011; Piwowar et al., 2018; Tenopir et al., 2017; Wallis, Rolando & Borgman, 2007). Todavia, devido à observação de fatores negativos em perspectivas individuais, institucionais e técnicas (Tabela 10), os investigadores parecem relutantes em adotar a partilha em acesso aberto (Borgman, 2012; Rowley et al., 2017).

As motivações, sejam elas positivas ou negativas, podem ser aproveitadas na RDM para a partilha, isto porque parece haver um consenso cada vez mais acentuado, justificado e direcionado principalmente pela ciência aberta, para que a partilha dos dados em acesso aberto seja adotada e realizada. Nesse sentido, pode-se trabalhar para minimizar as barreiras, de modo que as motivações para a partilha sejam postas em relevo, o que não significa que isto seja simples ou fácil, mas diante das transformações e exigências contemporâneas da ciência, se torna necessário. Assim, as motivações negativas, ligadas diretamente ou não às positivas, devem ser identificadas, de maneira que as tensões e conflitos gerados e minimizados concorram para reduzir os impedimentos para a partilha em acesso aberto. É também neste propósito que a RDM aqui objetivada deve considerar ambos os fatores como propulsores para que a suas práticas resultem na partilha dos dados em acesso aberto, permitindo que estas forças motivadoras sejam organizadas e passíveis de direcionamentos por atores articulados, para uso através de práticas essenciais que, aplicadas à RDM, proporcionam uma partilha adequada dos dados de investigação no ambiente da universidade.

6.5. CONCLUSÕES PARA A CONCEPÇÃO DO MODELO PARA A GESTÃO DOS DADOS

Os levantamentos qualitativos, resultados da análise da literatura e da discussão dos dados oriundos das entrevistas, são os principais elementos utilizados para a elaboração do modelo para a RDM que possibilita uma partilha adequada dos dados, no ambiente da universidade, com base na articulação dos atores e no uso de práticas essenciais para a gestão. Assim, foi apresentada a relação conceitual entre os eixos de estudo, bem como analisados e discutidos os dados empíricos. As conclusões destes levantamentos para a concepção do modelo estão sintetizados na tabela 23 e na tabela 24, que, no caso do nexos teórico, são as definições conclusivas e orientadoras sobre os eixos de estudo, e, no caso da análise e discussão dos dados empíricos, são as principais conclusões e recomendações resultantes. Seguem primeiro as definições dos eixos de estudo e depois as principais conclusões das entrevistas.

As definições conclusivas e orientadoras sobre os eixos de estudo, estão caracterizadas nos itens identificados abaixo, centrais para a concepção do modelo para a RDM:

- 1) Uma partilha adequada dos dados de investigação;
- 2) As práticas essenciais a serem utilizadas na RDM e
- 3) A articulação dos atores para a RDM na universidade.

Tabela 23

As definições centrais do modelo para a RDM na universidade

Definição	Descrição
Partilha adequada dos dados de investigação	É um processo sistêmico e organizado para a abertura dos dados de investigação, realizada por um conjunto de atores associados, centrada no investigador como ator principal, sob o qual devem ser observadas as influências cognitivas, contextuais e sociotécnicas. Fundamentada no uso relacional e intencional dos princípios e práticas da ciência aberta, o processo para a partilha dos dados deve ser suficientemente flexível para abarcar as especificidades das áreas de conhecimento, as orientações e as exigências das políticas, bem como as documentações que facilitam o uso atual e futuro dos dados, sendo assegurada por infraestruturas digitais e sempre que possível na forma do acesso aberto, com todo o processo objetivado para a reutilização dos dados de investigação.
Práticas essenciais para a RDM na universidade	É um conjunto de ações para gerir e partilhar os dados de investigação de forma adequada, sendo constituído por um fluxo de etapas sequenciais iniciadas com o planeamento e realizadas através da recolha e análise dos dados, tendo a sua finalização com o ato da partilha. O fluxo é permeado por práticas contextuais que abrangem duas ou mais etapas sequenciais, de modo a garantir a descrição e a monitoração dos dados, mas também a constituir a proteção legal e técnica necessária, bem como os recursos estruturais para preservação, gestão e a partilha dos dados de investigação. Sendo direcionadas ao acesso aberto, sob os princípios da ciência aberta e objetivada para a reutilização dos dados, as práticas essenciais devem ser associadas ao ambiente e ao conhecimento contextual, proporcionando também a distinção entre as práticas da RDM e da curadoria digital, de modo a evitar a justaposição de trabalhos.
Articulação dos atores para a RDM na universidade	É um tipo de articulação sistêmica que possibilita a realização de práticas essenciais para a RDM e uma partilha adequada dos dados, através do relacionamento entre os elementos humanos e não humanos, constituídos a partir dos componentes envolvidos para a RDM na universidade, sob o acesso aberto e fundamentado na ciência aberta. Para tanto, os atores são agrupados por classes de investigadores, comunidades, ferramentas, políticas e atribuições para o trabalho, viabilizados por um método interventivo de aprendizagem que permite a adaptação ao ambiente, mediante um processo onde os atores são sistematizados, os problemas enquadrados e avaliados para conceber soluções, que após modeladas e postas em testes podem ser efetivadas para o uso, ou retornar ao processo para aperfeiçoamento. Quando efetivado, o processo resulta em atores articulados e uma RDM adaptada ao ambiente, para a realização de práticas essenciais de gestão e uma partilha adequada dos dados de investigação.

Na tabela abaixo, as principais conclusões da análise e da discussão dos dados coletados por meio do levantamento empírico, estão representadas por um conjunto de sentenças decorrentes agrupadas pelos blocos temáticos abordados.

Tabela 24

Conclusões e recomendações oriundas dos dados empíricos

Tema	Sentido das proposições expressas acerca da RDM na universidade, obtidas da análise e discussão dos dados empíricos
Ambiente da investigação	1. A compreensão das especificidades das investigações, no ambiente diverso da universidade, é fundamental para poder gerir e partilhar os dados.
	2. O modelo resultante deve possibilitar um viés generalista do ambiente, sem deixar de considerar as especificidades e as diferenças disciplinares.
	3. O investigador é o ator principal e deve ser observado sob a influência do ambiente que repercute nas práticas para a gestão e partilha dos dados.
	4. O ambiente das UI&D onde a investigação está alocada, deve ser compreendido, pois remete às dificuldades e às diferenças entre as UI&D.
	5. A minimização das dificuldades e das diferenças no ambiente são medidas necessárias para a RDM na universidade.
	6. A compreensão do ambiente deve abranger as investigações e as UI&D, em uma abordagem multidimensional entre o individual e o institucional.
Caracterização dos dados	7. Os dados gerados no ambiente necessitam ser caracterizados, estudados e organizados para perceber o valor que possuem, como e onde armazená-los, o que fazer com eles e os danos que suas perdas podem causar, bem como evidenciar os benefícios decorrentes de uma correta gestão.
	8. A caracterização dos dados deve ser acompanhada por uma leitura do ambiente da investigação para que possam ser contextualizados.
	9. Para gerir volumes, tipos e formatos de dados heterogêneos, a base de suporte pode ser gradualmente desenvolvida, alcançando os itens comuns e mais simples, seguindo para uma maior complexidade e abrangência, de modo a permitir as articulações e práticas para a partilha dos dados.
	10. Um ator primordial é o investigador, como detentor de informações essenciais, para sugerir onde os dados podem ou poderiam ser reutilizados.
	11. O uso ou reuso dos dados de forma interdisciplinar ou pela sociedade, precisa de informações representativas para que sejam objetos informacionais relevantes.
	12. Para que a reutilização dos dados seja pertinente, é preciso que as práticas para a RDM e para a partilha sejam realizadas com esse objetivo, sendo necessário uma estrutura de organização e apoio, de modo a oferecer suporte para que o investigador seja direcionado a gerir os dados para a reutilização.
	13. A RDM deve pôr em relevância o valor dos dados, com práticas de proteção e preservação que possam garantir o conhecimento e a utilização deste valor na atualidade e em longo prazo. Ou seja, o valor que os dados têm ou que ainda podem possuir, deve estar presente na RDM.
	14. Deve-se construir uma estrutura de atenção para que o valor dos dados seja observado com a devida importância, através de práticas que os identifiquem e os façam conhecidos publicamente dentro e fora da universidade, realizada através de uma RDM abrangente no ambiente.
	15. É requerido do investigador que atente com instrumentos necessários para o nível de sensibilidade dos dados, sendo preciso analisar e determinar de modo nítido o que é público e o que é privado, o que é pessoal, confidencial ou sensível.

	<p>16. Os dados sensíveis devem ser vistos sob a possibilidade de poderem, ou não, serem partilhados, um cuidado que deve permear a RDM, mas também podem gerar conflitos que interferem na realização da partilha.</p> <p>17. Os instrumentos éticos, as ferramentas tecnológicas e políticas bem definidas, devem ser utilizadas em relação aos dados de natureza sensível, juntamente com as práticas para constituir a proteção legal e técnica, para evitar uma partilha inadequada e permitir a segurança para partilhar os dados necessários.</p> <p>18. A proteção dos dados deve ser realizada de modo claro, posta em prática através de instrumentos que possam ser utilizados para apoio ao investigador, mas também de acordo com as especificidades dos dados.</p>
Pessoas responsáveis	<p>19. É preciso identificar as áreas de responsabilidades e definir as atribuições para as pessoas participantes da RDM, de modo formal e, se possível, compondo a estrutura organizacional, com o objetivo de normatizar o apoio ao investigador para a RDM.</p> <p>20. Deve haver um planeamento e uma divisão do trabalho em relação as pessoas envolvidas na RDM, considerando as práticas atualizadas das tecnologias informáticas e o contexto ambiental, de modo que permita soluções adequadas à realidade da investigação e das UI&D, evitando a justaposição de responsabilidades e a sobrecarga de trabalho.</p> <p>21. O investigador deve participar do planeamento das atribuições e responsabilidades das pessoas envolvidas na RDM, pois é o investigador quem lida diretamente com os dados e com os vários aspectos para a sua gestão e partilha.</p>
Processos e fluxos de trabalho	<p>22. O processo ou fluxo deve ser flexível e adaptado às necessidades e ao perfil da investigação, de modo que contenha práticas comuns, mas que também permita a inserção de novos elementos para a organização do trabalho.</p> <p>23. Deve-se tratar os fluxos de forma atenta ao processo de investigação e ao comportamento dos participantes, para obter o mínimo de perdas informacionais.</p> <p>24. Um processo ou fluxo de trabalho sob a filosofia aberta pode ser pensado no sentido da construção comunitária e colaborativa.</p> <p>25. A preservação dos dados pode ser direcionada pelos fluxos e processos de trabalho, minimizando as perdas por obsolescência ou desconhecimento.</p>
Colaborações e influências	<p>26. As práticas e a articulação dos atores para a RDM devem canalizar esforços para uma colaboração positiva e organizada na partilha dos dados.</p> <p>27. É necessário estabelecer de uma rede colaborativa de articulações abrangentes para explorar o potencial comum, com a participação dos gestores e da liderança académica da universidade.</p> <p>28. Apresenta-se a necessidade de atores humanos para auxiliar os investigadores em questões que envolvam a RDM no ambiente.</p> <p>29. A literacia, a promoção e o treinamento na educação aberta é importante para suprir lacunas no conhecimento e habilidades em relação à RDM.</p> <p>30. A transversalidade da ciência aberta, dos recursos económicos e da tradição disciplinar, são fatores de atenção que podem influenciar em aspectos positivos e negativos o investigador para que as colaborações sejam direcionadas em benefício da RDM.</p> <p>31. É preciso atentar para atualizações e práticas colaborativas no ambiente “aberto”, permitindo adaptar-se a processos e tecnologias sob a filosofia aberta.</p> <p>32. É necessário definir estratégias para que os recursos económicos sejam administradas para o constantemente desenvolvimento da RDM.</p>

Ferramentas e serviços	33. O uso de instrumentos, tecnologias e serviços para o apoio da RDM na universidade, direcionados pelos meios digitais, exercem participação essencial para desenvolver e articular atores e práticas para a gerir e partilhar os dados de investigação	
	34. O correto dimensionamento da infraestrutura organizacional e tecnológica deve ser pensada de modo a melhor aproveitar as ferramentas e serviços, com o uso de propriedades intrínsecas das redes colaborativas e da multidisciplinaridade.	
	35. É necessário que as ferramentas e serviços sejam trabalhadas juntamente com outras competências, de modo que a RDM seja realizada sob uma infraestrutura não somente ferramental e tecnológica, mas esclarecedora da importância relacional para os seus usuários.	
	36. As infraestruturas devem ser observadas sob as características de um conhecimento comum, com recursos partilhados, observadas sob a perspectiva de que as pessoas estão sujeitas a dilemas sociais e sob tensões abundantes entre as partes interessadas.	
	37. A interoperabilidade entre os materiais que produzem e gerenciam os dados precisa ser desenvolvida para que o aspecto colaborativo da RDM possa ser maximizado.	
	38. É necessário pensar a interoperabilidade em múltiplos sistemas e ferramentas de informação no âmbito da instituição e fora dela.	
	39. A interoperabilidade deve ser realizado no sentido <i>bottom-up</i> e <i>top-down</i> , de modo a envolver a universidade com múltiplos olhares em níveis gerais, juntamente com as especificidades da investigação.	
	40. As práticas sustentáveis devem estar diretamente ligadas às ferramentas e serviços, juntamente com ações de governança, para que sejam sustentáveis e proporcionem a sustentabilidade.	
	41. O planeamento para delinear como os dados serão armazenados e como o acesso e a segurança serão gerenciados, deve ser uma prática para a organização e minimização do risco de perdas informacionais.	
	42. É pertinente envolver equipes técnicas de várias áreas juntamente com a governança da universidade, em um arranjo lógico e sob igual importância, para abordar questões amplas como as estruturas a nível da universidade, ou mais específicas como a usabilidade dos materiais.	
	43. É preciso dimensionar e articular as infraestruturas tecnológicas e organizacionais, além de determinar o grau de investimento na preservação em longo prazo, com uso e construção de políticas, estratégias e infraestruturas direcionadas para os cuidados com os dados	
	44. A infraestrutura para a RDM necessita estar interligada ao ambiente tecnológico e computacional.	
	45. Precisa-se que a governança da universidade utilize investimentos financeiros, políticos e de capacitação para tornar a infraestrutura articulada interoperável e sustentável, mas também para minorar tensões e conflitos decorrentes.	
	adotadas	46. As políticas, regras e documentos existentes ou a serem elaborados que envolvem a RDM, devem ser prioridade para as práticas que possibilitam a gestão e a partilha dos dados na universidade.
		47. As políticas devem ser organizadas e implementadas para funcionar como um agregador de princípios e práticas que auxiliam os investigadores a lidar com os dados gerados.
48. As questões legais devem ser organizadas nos níveis global e local, de modo que sejam complementares e não sobrepostas.		

Políticas, normas e orientações	49. A construção de uma proteção legal sob a RDM, deve ser direcionada para diminuir as vulnerabilidades e proporcionar segurança no contexto da investigação e dos seus atores.
	50. Os metadados são obrigatórios como uma prática para assegurar que a RDM possa ser realizada para tornar os dados reutilizáveis.
	51. É necessário dispor das condições técnicas e infraestruturais, de modo que o investigador possa ser apoiado por meios articulados e que os metadados sejam criados a contento das especificidades da investigação.
	52. A biblioteca é um atores importante que pode colaborar para a RDM em termos de documentação de padrões e metadados.
	53. As UI&D, as investigações e investigadores devem ser preparados através de políticas, infraestrutura, esclarecimentos e práticas para que a elaboração de um plano de gestão de dados seja recebido de modo a gerar o mínimo de barreiras.
A partilha dos dados em acesso aberto	54. A ação da partilha requer uma série de práticas até chegar a este ato, de forma a garantir os benefícios que culminam da reutilização dos dados.
	55. O ato da partilha deve ser visto em conjunto e em contexto com o ambiente da investigação e do investigador.
	56. Há necessidade em saber onde estão os dados já partilhados.
	57. Deve-se atentar aos motivos da não partilha dos dados, como um trabalho volumoso e que requer o envolvimento dos atores para a minimizá-los.
	58. É necessário atentar, de modo organizado e articulado, as motivações como elementos propulsores para a partilha dos dados em acesso aberto.
	59. Deve-se trabalhar para minimizar as barreiras, de modo que as motivações negativas sejam identificadas e as positivas sejam postas em relevo.
	60. As motivações negativas e positivas não podem ser vistas isoladamente, pois estão intimamente relacionadas entre si e ao contexto.
61. As soluções devem ser pensadas de modo global a partir de práticas locais.	

7. PROPOSTA DE MODELO PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO

A proposição de um modelo para a RDM (*Research Data Management*), objetivo geral deste trabalho, tem na relação entre as abordagens teórica e empírica aqui realizadas o direcionamento necessário para subsidiar tal proposta, cuja finalidade está em promover uma partilha adequada dos dados de investigação, através da articulação dos atores e no uso de práticas essenciais para a RDM no ambiente da universidade. O direcionamento das abordagens foi responsável pela sinalização de conceitos, atores, elementos, processos, características e metodologia, entre outros aspectos constantes do resultado apresentado no modelo final. Este resultado é aqui evidenciado através de uma representação gráfica, acompanhada pela sua versão textual. Antes, porém, realiza-se uma breve caracterização sobre os modelos científicos, identificando as particularidades do modelo para a RDM proposto.

7.1. MODELOS CIENTÍFICOS

Os modelos têm uma relevância significativa no modo de fazer ciência (Hey, Tansley & Tolle, 2009) e podem ser classificados, caracterizados e definidos de muitas formas, em variados tipos, naturezas, funções, modos de expressões e características, sobre quase todas as áreas do conhecimento, com adaptações e readaptações para outras áreas, sendo possível encontrar trabalhos que envolvem discussões sobre modelos e modelagem nos mais diversos campos e propósitos (Da Silva, 2010; Frigg & Hartmann, 2020; Nouvel, 2002). Assim, para a definição de um modelo que sirva de base para a estrutura da RDM objetivada, considerando a pluralidade envolvente dos modelos, somado ao também diverso e complexo contexto do estudo abordado, optou-se como mais adequado não procurar uma classificação pronta para adaptar o modelo para a RDM, mas buscar caracterizações que o indique e o represente em relação às suas especificidades e ao que se propõe. Portanto, sob tais considerações, seguem abaixo alguns aspectos encontrados que podem retratar de forma geral o modelo para a RDM aqui proposto, através das seguintes características:

a) Representativo – como uma estrutura abstrata e simplificada de uma realidade, apresenta as relações sob uma forma aproximada, com a qual se pode explicar ou analisar o comportamento do que foi modelado, em um todo ou em partes, de modo a contribuir com

aquilo que se sabe e aquilo que se quer compreender ou simular (Capra, 1982; Cougo, 2013; Haggett & Chorley, 2013; Morrison & Morgan, 1999).

b) Heurístico – visa romper com uma interpretação inadequada para conceber uma interpretação nova e mais adequada, que será a nova linguagem através do qual o objeto pode ser apreendido como novidade possível de evolução, de modo a possibilitar a percepção de outros aspectos a serem considerados (Da Silva, 2010; Black, 1962).

c) Sistêmico – com a representação de processos e interações criadas com um objetivo específico, segundo a qual a realidade é vista em termos de sistemas e subsistemas interligados, dependentes e possíveis de serem analisados, de modo a evidenciar continuamente as características do ambiente, apontar as inter-relações e os sistemas subjacentes (Bertalanffy, 1973; Fiske, 2004; Justi, 2006).

d) Informacional – em relação à informação sistêmica, possibilita representar os elementos humanos e seus fatores cognitivos, como parte de um sistema de informação com fatores limitadores e elementos não humanos, bem como as relações de aquisição, organização e manipulação da informação também em uma interação homem-máquina (Burt & Kinnucan, 1990; Sayão, 2001).

e) Cognitivo – no sentido de que o modelo orienta a interpretação dos elementos humanos, sistêmicos e a interação entre eles, de modo a perceber o ser humano na realidade contextual, a identificação dos componentes dos modelos e dos seus elos, tendo como importante o modo do conhecimento que esses elementos têm de si próprios, de cada um dos outros componentes e da própria realidade contextual (Allen, 1991; Daniels, 1986).

f) Expansivo – permite que a partir da estrutura modelar sejam contempladas sugestões para a sua própria extensão, aplicações e reaplicações, por meio de operações e transformações proporcionadas e direcionadas por suas leis estruturais, oportunizando a mutabilidade e a dinâmica do contexto, de modo a permitir a otimização a partir do modelo (Black, 1962; Chorley & Haggett, 2013; Hesse, 1953; Jones et al., 2000)

e) Mediador – é um instrumento de mediação entre as teorias e o mundo, postos em uso no modelo e a ser aplicado de modo real, ao mesmo tempo que está inserido em um sistema lógico e gerador de conhecimento (Morrison & Morgan, 1999; Pietrocola, 1999).

f) Operacional – no sentido de não ser um instrumento passivo, mas direcionado por uma operacionalização teórica, através de uma intervenção prática em busca de resultados concretos, podendo ser considerado como uma tecnologia capaz de fornecer instrumentos de investigação que possibilitam a compreensão das teorias e do mundo (Da Silva, 2010; Morrison, 1999).

g) Exploratórios – não fornece uma descrição precisa da população real, mas possibilita aprender algo sobre um sistema específico, sendo utilizado como uma ferramenta para que, a partir do exame das características teóricas, forneça o ponto de partida para desenvolver modelos mais realistas (Fisher, 2006; Massimi, 2019).

m) Interdisciplinar - é o modelo que se inscreve em mais de uma categoria modelar, podendo ser utilizado em contextos investigativos e disciplinares diferentes, assumindo e considerando os aspectos cognitivos, físicos, sociais e comportamentais (Da Silva, 2010; Niedzwiedska, 2003).

7.2. REPRESENTAÇÃO GRÁFICA DO MODELO PARA A GESTÃO DOS DADOS

O modelo gráfico para a RDM na universidade, apresentado na figura 42, pode ser visualizado através das sete perspectivas que o compõem, enumeradas abaixo:

- 1) As Unidades de Investigação e Desenvolvimento (UI&D);
- 2) Os dados a serem geridos e abertos para a reutilização;
- 3) A partilha adequada dos dados de investigação;
- 4) As práticas essenciais para a gestão e a partilha dos dados de investigação;
- 5) A articulação dos atores para a gestão e a partilha dos dados de investigação;
- 6) As forças transversais influenciadoras; e
- 7) A metodologia para adaptação contextual.

As perspectivas que compõe a estrutura modelar são o resultado de cinco pontos orientadores baseados nos objetivos específicos deste trabalho, que permitiram através da sua conjugação a concepção do modelo para a RDM.

O primeiro ponto orientador indicou a necessidade de conhecer as características do objeto informacional a ser gerido – os dados de investigação (Tabela 8) –, aberto, partilhado e reutilizado. Este ponto foi resultado da contextualização teórica do ambiente, que direcionou a possibilidade de agrupamento dos atores para a RDM na universidade em classes genéricas, além da própria caracterização dos dados de investigação.

O segundo ponto orientador obteve na ciência aberta os parâmetros necessários (Tabela 14) para caracterizar uma partilha adequada dos dados de investigação (Tabela 23), também assinalou a existência de forças influenciadoras (Figura 6), pautando as forças que influenciam para a RDM e principalmente para a partilha dos dados, bem como proporcionou o contexto e os elementos basilares para que os dados sejam partilhados (Figura 8).

O terceiro ponto são os elementos e os processos presentes nos modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, indicando as práticas essenciais para a RDM (Figura 25) que levam sustentação para uma partilha adequada dos dados, mas também obteve as contribuições para o desenvolvimento do modelo (Tabelas 19 e 20). A nomenclatura original e a ordem das etapas encontradas para as práticas essenciais foram adaptadas no modelo final, porém sem alteração da sua essência.

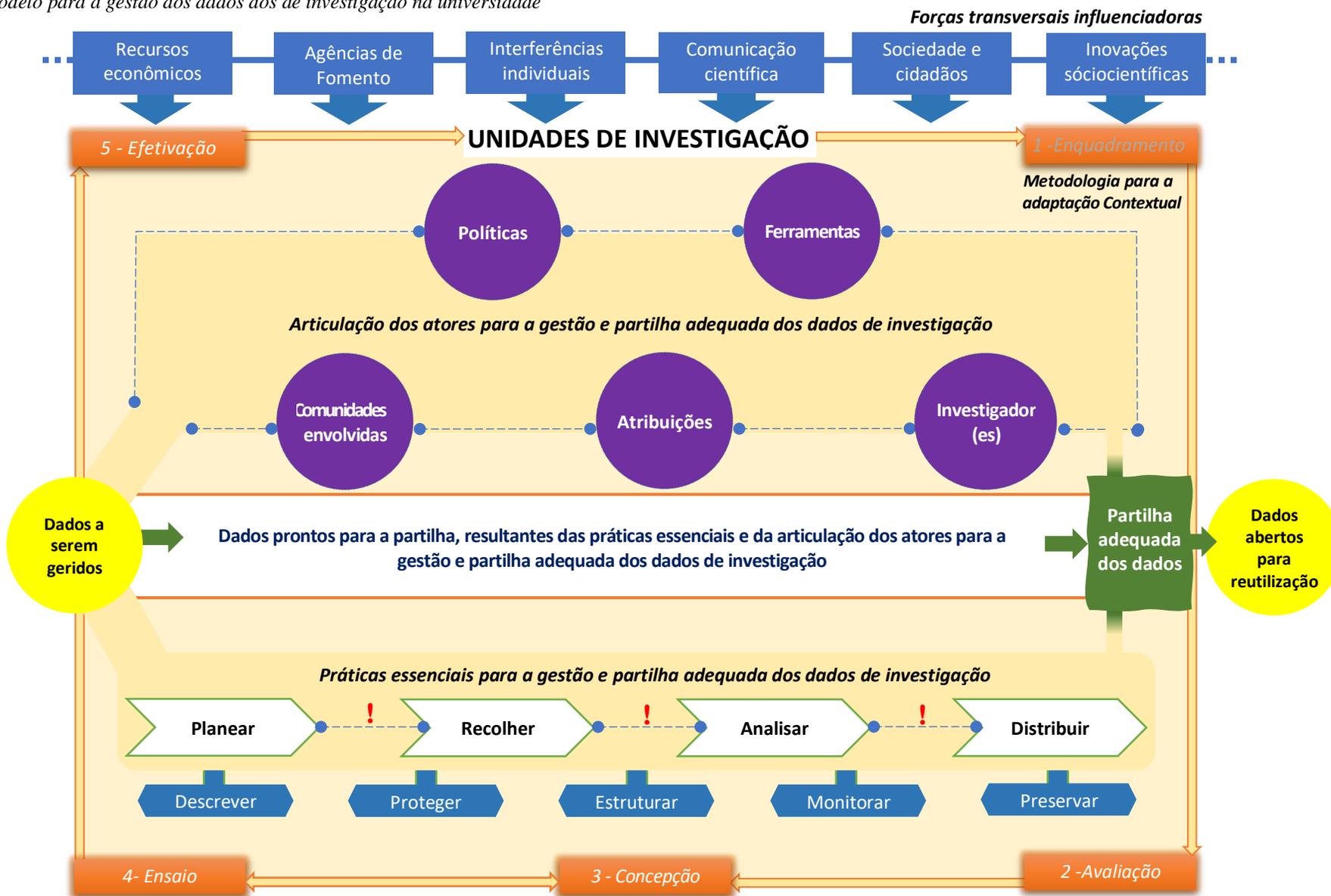
O quarto ponto orientou a articulação dos atores para a gestão e partilha dos dados, no desenvolvimento da metodologia para adaptação contextual (Figura 31), originados na teoria da atividade e ajustados aqui para uso no modelo final. Este ponto indicou como sistematizar e atentar para as especificidades do ambiente, de modo a proporcionar que a RDM na universidade possa ser realizada através das articulações dos atores e das práticas essenciais de gestão, objetivadas em proporcionar uma partilha adequada dos dados.

O quinto ponto norteador concerne aos aspectos provenientes da coleta dos dados empíricos que, em contraste com a teoria, permitiu subsidiar todas as perspectivas do modelo final para a RDM (Tabela 23), indicando também duas forças transversais influenciadoras, quais sejam: os recursos econômicos e as inovações sociocientíficas.

A figura a seguir apresenta a representação gráfica do modelo final para a RDM na universidade, posteriormente é também apresentado a sua versão textual.

Figura 42

Modelo para a gestão dos dados de investigação na universidade



7.3. MODO TEXTUAL DO MODELO

A versão em texto do modelo é a explicação dos seus elementos, fluxos e relações, derivados dos resultados das estratégias de investigação empreendidas ao longo deste estudo. Assim, para um melhor entendimento e interpretação, o modelo gráfico para a RDM na universidade (Figura 42) será explicado através de quatro conceitos centrais, depois por uma visão geral das sete perspectivas que o constitui (Figura 43), seguido de um panorama mais detalhado de cada uma das perspectivas do modelo para a RDM.

7.3.1. OS CONCEITOS CENTRAIS

Derivados do estudo dos eixos teóricos, os conceitos centrais que orientam as perspectivas do modelo são as definições conclusivas resultantes da análise da literatura direcionadas aos objetivos desta tese. Portanto, seguem os quatro conceitos centrais decorrentes e que norteiam o modelo para a RDM apresentado:

A partilha adequada dos dados de investigação: é um processo sistêmico e organizado para a abertura dos dados de investigação, realizada por um conjunto de atores associados, centrada no investigador como ator principal, sob o qual devem ser observadas as influências cognitivas, contextuais e sociotécnicas. Fundamentada no uso relacional e intencional dos princípios e práticas da ciência aberta, o processo para a partilha dos dados deve ser suficientemente flexível para abarcar as especificidades das áreas de conhecimento, as orientações e as exigências das políticas, bem como as documentações que facilitam o uso atual e futuro dos dados, sendo assegurada por infraestruturas digitais e sempre que possível na forma do acesso aberto, com todo o processo objetivado para a reutilização dos dados de investigação.

As práticas essenciais para a RDM: é um conjunto de ações para gerir e partilhar os dados de investigação de forma adequada, sendo constituído por um fluxo de etapas sequenciais iniciadas com o planeamento e realizadas através da recolha e análise dos dados, tendo a sua finalização com o ato da partilha. O fluxo é permeado por práticas contextuais que abrangem duas ou mais etapas sequenciais, de modo a garantir a descrição e a monitoração dos dados, mas também a constituir a proteção legal e técnica necessária, bem como os recursos estruturais para preservação, gestão e a partilha dos dados de investigação.

Sendo direcionadas ao acesso aberto, sob os princípios da ciência aberta e objetivada para a reutilização dos dados, as práticas essenciais devem ser associadas ao ambiente e ao conhecimento contextual, proporcionando também a distinção entre as práticas da RDM e da curadoria digital, de modo a evitar a justaposição de trabalhos.

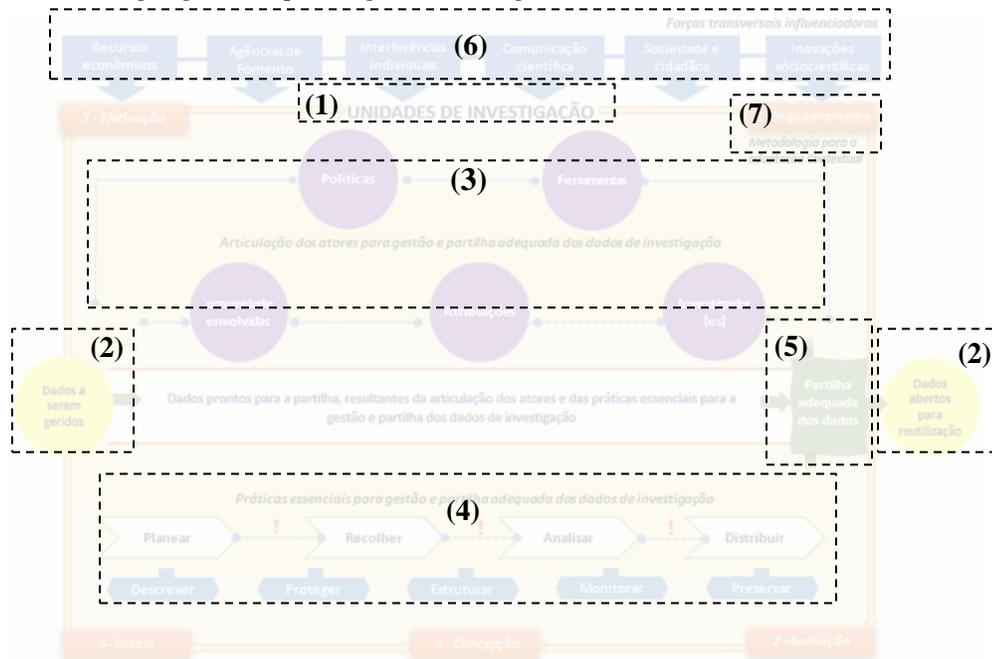
A articulação dos atores para a RDM: é um tipo de articulação sistêmica que possibilita a realização de práticas essenciais para a RDM e uma partilha adequada dos dados, através do relacionamento entre os elementos humanos e não humanos, constituídos a partir dos componentes envolvidos para a RDM na universidade, sob o acesso aberto e fundamentado na ciência aberta. Para tanto, os atores são agrupados por classes de investigadores, comunidades, ferramentas, políticas e atribuições para o trabalho, viabilizados por um método interventivo de aprendizagem que permite a adaptação ao ambiente, mediante um processo onde os atores são sistematizados, os problemas enquadrados e avaliados para conceber soluções, que após modeladas e postas em testes podem ser efetivadas para o uso, ou retornar ao processo para aperfeiçoamento. Quando efetivado, o processo resulta em atores articulados e uma RDM adaptada ao ambiente, para a realização de práticas essenciais de gestão e uma partilha adequada dos dados de investigação.

A Reutilização dos dados de investigação: garante que os dados estejam acessíveis após o término da investigação, tanto para a comunidade científica quanto para qualquer usuário interessado em reutilizar os dados, de modo a envolver e repercutir na ciência e na sociedade. A reutilização é resultado da articulação dos atores e das práticas essenciais para a RDM, que produz uma partilha adequada dos dados, para que os dados partilhados possam também ser úteis dentro de uma estrutura conceitual diferente da qual foram concebidos, permitindo a geração de novas ideias para as investigações atuais e futuras. Todos os elementos do modelo articulam-se e realizam práticas objetivadas para que os dados sejam reutilizados através de uma partilha adequada.

7.3.2. A VISÃO GERAL

Figura 43

Visão geral das sete perspectivas que compõe o modelo para a RDM

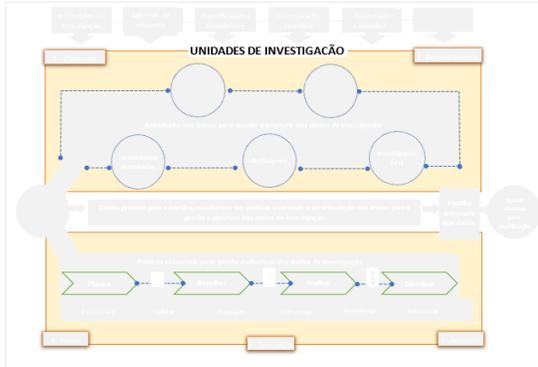


O ambiente para aplicação do modelo para a RDM está definido na perspectiva para uso nas **unidades de investigação (1)**, que contém o ambiente da investigação e uma ligação institucional com a universidade. Assim, **os dados de investigação precisam ser geridos (2)** em duas perspectivas: através da **articulação dos atores (3)** e através de **práticas essenciais (4)**, ambas objetivadas para uma gestão e partilha adequada dos dados. Estas duas perspectivas podem ser assim efetuadas porque estão **ligadas e direcionadas para a partilha adequada dos dados (5)**. A articulação dos atores refere-se a um nível global mais institucional e as práticas essenciais referem-se a um nível local mais específico da investigação. Ambas sofrem as **influências de forças transversais (6)** que interferem e devem ser consideradas para a realização da partilha dos dados. No entanto, para que o modelo para a RDM seja aplicado e todas as perspectivas sejam consideradas em conjunto, de forma organizada, a perspectiva da **metodologia para a adaptação contextual (7)** deve ser realizada, pois esta, por meio de uma construção multivocal, analisa as perspectivas do modelo e propõe uma forma para que a RDM seja adaptada ao ambiente, permitindo que através da conjunção de todos os elementos, sejam enviados **dados** prontos para uma partilha adequada, resultando em dados **abertos** e disponíveis **para a reutilização (2)**.

7.3.3. O AMBIENTE: AS UNIDADES DE INVESTIGAÇÃO DA UNIVERSIDADE

Figura 44

As unidades de investigação para a RDM



O modelo para a RDM proposto tem como gene ambiental as UI&D da universidade (representadas no modelo gráfico pela cor salmão). As unidades de investigação são aqui caracterizadas como o espaço de investigação onde os dados podem ser geridos em dois sentidos – o local e o global. Ou seja, a gestão

dos dados utilizados nas investigações das UI&D é realizada pela articulação dos atores (global) e por intermédio de práticas essenciais (local) para uma partilha adequada dos dados.

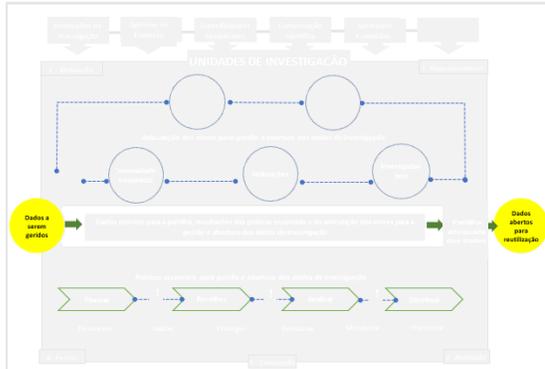
No sentido global, as UI&D estão em articulação direta com os atores, são os investigadores, mas também as políticas, as ferramentas, as atribuições de trabalho e com as comunidades e partes interessadas na RDM, em uma relação de interdependência de modo institucional. No sentido local, a perspectiva das práticas essenciais para a RDM ocorre em seu âmbito, através de ações que têm como eixo o processo de investigação e a curadoria dos dados a partir das investigações locais. O ambiente das UI&D estão permeadas por uma terceira perspectiva, as forças influenciadoras que podem interferir na RDM, impedindo ou colaborando no alcance do objetivo de partilhar os dados adequadamente.

A conjugação desses elementos e perspectivas observadas no ambiente permite que as UI&D, contenham em si tanto uma articulação realizada por um conjunto de atores associados, quanto um fluxo de práticas essenciais ambientadas sob o processo da investigação. No entanto, como os espaços de investigação possuem características diferentes uns dos outros, tais espaços devem ser analisados sob uma quarta perspectiva, a da metodologia para a adaptação contextual, de modo que a RDM possa ser adaptada ao ambiente e às especificidades da investigação. Interrelacionadas, as perspectivas resultam em dados prontos para partilha e disponíveis para a reutilização. Portanto, no modelo proposto, as UI&D constituem-se como o ambiente em que a RDM pode ligar os aspectos global e local, com atores articulados e na realização de práticas essenciais para a gestão e a partilha adequada dos dados.

7.3.4. OS INSUMOS BÁSICOS: OS DADOS DE INVESTIGAÇÃO

Figura 45

Os dados de investigação para a RDM no modelo



Os dados brutos da investigação, no ambiente das UI&D, são os insumos básicos de entrada do modelo para a RDM, a serem transformados em dados abertos e disponíveis de forma pública para uso e reuso. Ou seja, é o objeto informacional inicial que após gerido, é transformado em produto possível de ser

partilhado em acesso aberto e reutilizável (a entrada e a saída dos dados estão ilustradas no modelo pela cor amarela em formato circular). A fluidez dos dados de investigação para a RDM no modelo, segue na seguinte forma: os dados a serem geridos pelas UI&D constituem o objeto informacional principal, tanto da articulação dos atores, quanto das práticas essenciais, sendo, ao mesmo tempo, o alvo de atenção destas duas perspectivas cujo resultado são dados prontos para a realização de uma partilha adequada. Assim, os dados brutos a serem geridos são transformados em dados abertos e disponíveis para reutilização.

Os dados de investigação na perspectiva das práticas essenciais para a RDM, são cuidados por indivíduos responsáveis por garantir que eles sejam partilháveis em acesso aberto e reutilizáveis, assegurando que, durante o processo de investigação, a caracterização e a informação sobre os dados sejam realizadas, do mesmo modo que garantem a sua proteção e preservação. As práticas essenciais também asseguram que os dados sejam sustentados por estruturas planeadas e que haja o monitoramento para certificar que uma partilha adequada possa ser realizada. Na perspectiva da articulação dos atores para a RDM, os investigadores têm os dados da investigação amparados pelas políticas e pelas ferramentas estruturadas, observada também a relação destes dados com as comunidades envolvidas, bem como as atribuições de trabalho que permitem a realização da RDM através das responsabilidades pessoais.

As outras duas perspectivas do modelo agem indiretamente sobre os dados, por ocasião da influência das forças transversais na RDM e na partilha dos dados, e quando, por

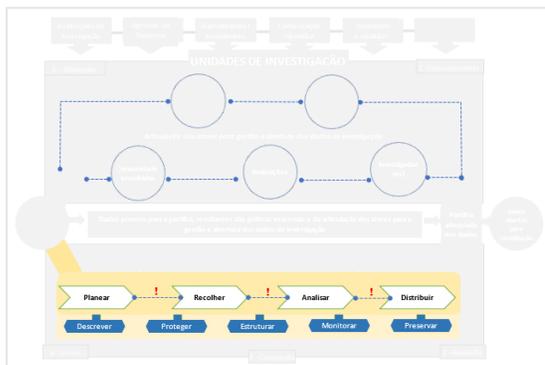
exemplo, a metodologia para adaptação contextual levanta as especificidades do contexto, que incluem a caracterização dos dados para o ambiente da investigação.

Portanto, os dados de investigação objetivados para uma partilha adequada são os objetos informacionais fundamentais de duas perspectivas do modelo, de modo concomitante, relacional e com igual valor de importância. Uma perspectiva, que pode ser vista no sentido global, é constituída por atores articulados e responsáveis pelos aspectos estruturantes para a gestão e o sustento dos dados. A outra perspectiva, vista no sentido local, é constituída por práticas essenciais para a gestão e curadoria dos dados. Significa dizer que os dados que alimentam e resultam das investigações das UI&D têm a sua gestão direcionada para uma partilha adequada dos dados de investigação, promovidas primordialmente pela articulação de atores e por práticas essenciais de gestão, de modo a transformar os dados brutos, a serem geridos, em dados abertos e prontos para a reutilização.

7.3.5. AS PRÁTICAS ESSENCIAIS: GESTÃO E PARTILHA DOS DADOS

Figura 46

As práticas essenciais para a RDM no modelo



As práticas essenciais para uma gestão e partilha adequada dos dados, no âmbito das investigações nas UI&D, corresponde a um fluxo formado por etapas, realizadas a partir do processo de investigação e representado por um conjunto de práticas interconectadas e interdependentes. Primeiramente estruturada

em um fluxo de práticas sequenciais (identificado na cor branca em formato de seta larga), destinadas a distribuir os dados para uma partilha adequada. Entre uma etapa sequencial e outra, uma seta de exclamação (identificada na cor vermelha) representa a atenção com as perdas dos dados quando da passagem de uma etapa para a seguinte. Interligadas ao fluxo sequencial, estão as práticas contextuais (identificadas na cor azul clara), a interligação indica que as fases contextuais podem ser praticadas em qualquer etapa do fluxo sequencial, quando assim for pertinente. A aplicação em conjunto das etapas sequenciais e contextuais, resultam nas práticas essenciais para a gestão e a partilha adequada dos dados de investigação.

Assim, para gerir e partilhar os dados de investigação no sentido acima exposto, o modelo assume a perspectiva das práticas essenciais na conjugação de dois aspectos: as etapas sequenciais e as etapas contextuais. O primeiro aspecto, mais direcionado ao processo de investigação, compreende as práticas para planejar, recolher, analisar e distribuir os dados e quando aplicável, os materiais subjacentes. O segundo aspecto, mais concentrado na curadoria digital, possui as práticas para descrever, proteger, estruturar, monitorar e preservar os dados e, quando cabível, os materiais subjacentes. Com base nesses dois aspectos que formam a perspectiva das práticas essenciais, segue sucintamente um exemplo explicativo para o uso e a interpretação de cada etapa, bem como podem estar relacionadas com os atores articulados do modelo gráfico. O formato de uma exemplificação apenas, foi escolhida de modo a não limitar a utilidade do modelo e indicar que existem outros modos de utilização para a estrutura apresentada. Nesse sentido, primeiro explica-se o fluxo dos aspectos das práticas sequenciais e depois os aspectos das práticas contextuais.

Planear: é a prática realizada por quem tem a atribuição de projetar a investigação e a gestão dos dados, desempenhado necessariamente por *investigadores* participantes da investigação, mas não somente, pois deve-se também ser direcionado a um *conjunto de pessoas interessadas* que podem colaborar com o planeamento. Entre as possibilidades de planeamento destaca-se a concepção de um plano de gestão dos dados, com as projeções para a geração, coleta e criação dos dados, as *políticas* a serem adotadas, as *ferramentas*, serviços e as *estruturas* a serem utilizadas, a definição das *atribuições* e quaisquer outros recursos necessários para gerir, cuidar e partilhar os dados. Esta etapa também realiza práticas de curadoria com a previsão da *descrição dos dados* e os detalhes que envolvem a sua *proteção* e a *preservação*, ainda projetando um plano de *monitoramento* durante a investigação para verificar se os propósitos da curadoria estão sendo cumpridos. A falta de um plano ou a sua má elaboração, pressupõe a *perda de dados*. Portanto, a RDM começa com um planeamento antes do início da investigação e precedendo a prática para a recolha dos dados.

Recolher: esta etapa centraliza-se na coleta, captura ou geração dos dados brutos a serem utilizados na investigação, de acordo com o *planeamento* e as especificidades da investigação. São os dados primários e secundários obtidos através das *estruturas*, com o uso de *ferramentas* e equipamentos manuais ou automáticos, podem ainda ser oriundos de

indivíduos ou *comunidades envolvidas* com a investigação. A responsabilidade pela coleta e organização dos dados normalmente é dividida entre os *investigadores* da equipe. No entanto, podem existir *atribuições* deste trabalho partilhados com outros indivíduos, de modo a atentar para uma monitoria bem definida a fim de evitar as *perdas de dados* no processo da recolha, mas também para que fatores fundamentais, como a caracterização dos dados e suas conseqüentes *descrição, proteção e preservação* possam ser iniciadas já nesta etapa. Para tal, as *políticas* devem ser observadas e o *monitoramento* direcionado para que a coleta e o cuidado com os dados estejam conforme projetado. Esta etapa baseia-se no planeamento, para que a recolha e a organização dos dados brutos possam seguir para a etapa de análise.

Analisar: os dados brutos recolhidos e os dados processados na investigação são os alvos deste tipo de análise. As ações de análise nesta etapa do modelo são direcionadas para a gestão e a partilha dos dados, consistem principalmente na exploração, diagnóstico e percepção dos dados para que as práticas de curadoria os tornem compatíveis e reproduzíveis, conseqüentemente reutilizáveis. As ações da análise, por estarem sobrepostas ao processo da investigação, são *atribuídas* aos *investigadores* ou, no caso de apoio externo, com a sua participação direta. Os dados são aqui analisados, tendo como orientação o *planeamento* inicial e os requisitos exigidos pelas *políticas* e pela *comunidade* científica. As *estruturas* tecnológicas e o uso de *ferramentas* apropriadas devem assegurar que os dados sejam analisados conforme foram caracterizados na *descrição*, mas também sobre a possibilidade de descarte ou transformação dentro do contexto da *proteção e preservação* dos dados. Aqui o *monitoramento* deve garantir que o resultado da análise seja alvo de ação das demais fases e na minimização da *perda dos dados*. Todas as ações desta etapa são análises realizadas nos dados recolhidos e direcionadas para a sua distribuição.

Distribuir: significa concentrar o fluxo para culminar em uma partilha dos dados de forma pública e geral, também em termos de comunicação científica. Esta prática deve ser *monitorada* para que, antes da realização do ato, seja verificado se a distribuição dos dados está a ser realizada por *ferramentas* e *estruturas* compatíveis, com os dados aptos a serem *preservados*, resultando na minimização das *perdas* e na *proteção* a curto prazo e em longo prazo, conforme a direção do *planeamento* elaborado. A distribuição nesse sentido é realizada de acordo com as *políticas* para a partilha, constando as *descrições* dos dados por

metadados adequados e respeitando as considerações legais e técnicas adotadas, com o apoio dos recursos humanos, das *comunidades* e das instituições que podem estar envolvidas na demanda para a partilha. A chegada a esta última fase sequencial pressupõe que os investigadores e a investigação obtiveram as condições ideais para a distribuição dos dados, com os atores articulados e as demais práticas a funcionar de modo apropriado, para resultar em uma partilha adequada dos dados de investigação.

Descrito o fluxo sequencial, segue a explicação dos aspectos contextuais a ele interligado, formando assim as práticas essenciais para a RDM e a partilha dos dados.

Descrever: é a prática que atribui os metadados e os documentos necessários para a compreensão do contexto, interpretação, partilha e reutilização dos dados de investigação. São utilizados padrões apropriados para garantir uma descrição adequada e o uso em longo prazo, ou seja, refere-se a uma minuciosa e completa explicação sobre os dados e sobre o seu contexto, de modo a perceber também acerca do trabalho e da metodologia empregada. Assim, os padrões e *ferramentas* para a descrição dos dados devem estar em conformidade com as *políticas* e acordos nacionais e internacionais, bem como as exigências das *comunidades* envolvidas para sua interoperabilidade. Esta prática tem inserção em todas as fases do fluxo sequencial, cabendo aos *investigadores* especificar ou ratificar as *atribuições* e responsabilidades referentes a esta etapa, observado as especificidades da investigação.

Proteger: refere-se à segurança dos dados, dos indivíduos e das *comunidades envolvidas* durante e depois da investigação, abrange as ações para a realização de medidas de salvaguarda legal, direitos de propriedade, privacidade e confidencialidade, seja nas questões da proteção dos dados reservados, sensíveis ou pessoais, de interesses privados, públicos ou comunitários. Esta etapa também se preocupa com a proteção técnica e os aspectos de segurança dos sistemas informáticos, com o uso de *ferramentas* adequadas para as restrições de acesso, de modo a garantir a autenticidade e a confiabilidade dos dados e dos sistemas tecnológicos. As proteções devem ser utilizadas juntamente com as *políticas* gerais, observando as especificidades dos dados na investigação. Os *investigadores* são os responsáveis diretos por garantir esse tipo de proteção, mas pode haver *atribuições* específicas que o apoiem. Esta fase contempla ações em todas as etapas sequenciais.

Estruturar: nesta etapa realiza-se ações para garantir que a arquitetura, engenharia e as estruturas tecnológicas, juntamente com o *ferramental* necessário e específico para a investigação, sejam utilizadas de forma apropriada, ou seja, para que os dados sejam geridos com as mínimas dificuldades e de forma pertinente na investigação, permitindo o suporte necessário aos investigadores e a garantia do funcionamento das estruturas e infraestruturas físicas e tecnológicas, com o uso de plataformas de colaboração, comunicação, armazenamento e partilha dos dados, mas também para evitar a obsolescência do material. É suposto que as *atribuições* específicas para a realização desta prática recaiam sobre profissionais técnicos em trabalhos colaborativos com os investigadores. No entanto, os *investigadores* têm papel importante para determinar a estrutura, apoiado em *políticas* adequadas para prover as necessidades gerais e específicas para gerir e partilhar os dados da investigação. Nesta etapa, alvo de uso em toda as fases sequenciais, o monitoramento participa na interlocução com as *comunidades* envolvidas e com as instâncias superiores da universidade, tendo em vista as eventuais necessidades de solução para as dificuldades encontradas e os possíveis ajustes nas estruturas para a RDM e para a partilha dos dados.

Monitorar: são ações de gestão planeadas para promover de modo específico a curadoria dos dados. Nesse sentido, pode ser realizado por *investigadores* ou por pessoas com *atribuições* determinadas. As ações encontram-se no propósito de acompanhamento das estratégias, *políticas*, relacionamentos com instituições, entidades e *comunidades envolvidas*, mas também na preocupação com as *ferramentas* adequadas em relação a possíveis perdas de informações e de dados durante a investigação. Ou seja, o monitoramento tem o objetivo de acompanhar as práticas de curadoria no fluxo proposto, para verificar se estão a ser realizadas conforme planeadas e de modo adequado. Esta etapa articula com todas as outras etapas e com os gestores das UI&D, em conjunto com os gestores da universidade, para apoio e diálogo na resolução de eventuais problemas. É nesta etapa que se acompanham as ações para a abertura dos dados, bem como a preocupação com a utilização dos princípios FAIR (*Findable, Accessible, Interoperable and Reusable*), ambas monitoradas com o objetivo de gerir e partilhar os dados adequadamente.

Preservar: embora a preservação possa ser um tipo de proteção, aqui segue no sentido de garantir que os dados e todo o material necessário que o acompanha, quando

necessários, permaneçam autênticos e mantenham a sua integridade para a reutilização. Refere-se às práticas específicas que levam em consideração os aspectos para salvaguardar os dados, pelo tempo que for necessário, no curto prazo, para minimizar as perdas potenciais e, em longo prazo, para que as *comunidades* interessadas possam acessar, interpretar, usar e reutilizar os dados no presente e no futuro. Presume-se idealmente a designação específica de um *responsável* pelo monitoramento desta prática, não necessitando ser diretamente um dos *investigadores* envolvidos. A preservação dos dados deve ser preocupação em todas as etapas sequenciais. Tais considerações passam pela elaboração de planos e *políticas* de preservação e a efetiva guarda em locais com as *ferramentas* apropriadas às características dos dados e da investigação, mas também no cuidado com a recolha, organização e análise, de modo a manter o conteúdo dos dados e o seu contexto (descrição e metadados) seguros e preservados para a partilha e para a reutilização.

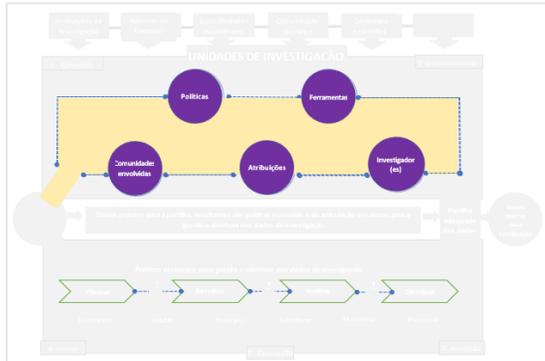
Perdas de dados! é o ponto de atenção na transição entre as etapas sequenciais, com a preocupação exclusiva na identificação das possíveis perdas de dados. Neste ponto, é papel das práticas de monitoramento estar a observar a possibilidade de tais perdas, como, por exemplo, assegurar que o método de seleção da amostra recolhida no processo de investigação seja incluído para a partilha dos dados. No entanto, isto não exime os *investigadores* ou outros atores envolvidos de estarem *atribuídos* de tal preocupação, podendo ser estabelecidas *políticas*, critérios, procedimentos e técnicas para verificar se durante as etapas sequenciais há imprecisões em relação ao tratamento dos dados, em busca de possíveis erros e proposições de soluções, de modo a minimizar as perdas e repercutir positivamente na gestão e na partilha adequada dos dados.

Como exposto nas explicações e para perceber o contexto das práticas essenciais para gerir e partilhar os dados de investigação, os investigadores têm particular responsabilidade na condução e na realização da perspectiva das práticas essenciais, sendo estas práticas também influentes na relação com todos os atores articulados do modelo. Por sua vez, a perspectiva dos atores articulados do modelo é necessária para o estabelecimento, a organização e o apoio para a realização das práticas essenciais, de modo que ambas as perspectivas dependem e são influenciadas reciprocamente.

7.3.6. A ARTICULAÇÃO DOS ATORES: GESTÃO E PARTILHA DOS DADOS

Figura 47

A articulação dos atores para a RDM no modelo



A articulação dos atores para a RDM (identificados no modelo na cor azul escura em formato circular) é realizada entre elementos complexos e díspares, com interesses distintos e interligados para alcançar a gestão e a partilha dos dados de investigação. Ou seja, os dados a serem

geridos passam por um conjunto de atores dependentes entre si que interagem institucionalmente, tendo como objeto a gestão para a partilha adequada dos dados.

São atores relacionais e organizacionais, de viés técnico e tecnológico, normativo e humano que, juntos e relacionados, são capazes de gerir os dados em direção ao seu objeto. Define-se aqui, como direção geral, que os gestores e os investigadores têm especial participação na orientação e no desenvolvimento das ações articulatórias. E que os atores articulados influenciam todas as etapas na perspectiva das práticas essenciais. Por sua vez, as práticas essenciais são necessárias para originar ou estabelecer a organização dos atores. Assim, ratifica-se que a articulação dos atores e a investigação na perspectiva das práticas essenciais são influenciadas reciprocamente. Os atores e as suas articulações são destacados a seguir.

Ferramentas: aglutina em si um conjunto de objetos fundamentais para realizar as práticas essenciais par a RDM, exercendo participação necessária para desenvolver e articular os demais atores. Tendo por base o aspecto técnico e material, este ator envolve diversos recursos e serviços que devem ser adaptados à realidade das investigações. São instrumentos físicos e lógicos, a serem geridos de modo institucional e local, que abrangem muitos meios e serviços. A organização deste ator de um modo lógico e coordenado, requer recursos direcionados para o seu desenvolvimento e manutenção, com o suporte estruturado das *políticas* e dos recursos humanos qualificados, *atribuídos* não apenas para a realização da parte técnica, mas também na interlocução com as *comunidades envolvidas* e com os demais *investigadores* para a RDM e na partilha adequada dos dados.

Políticas: este ator funciona para estruturar e apoiar as normas, regras, diretrizes e outros instrumentos explícitos e implícitos que sejam eficazes para a RDM, de modo institucional e também sob as especificidades da investigação. É um ator mandatário e orientador que influi diretamente na articulação dos demais atores e nas práticas essenciais para a RDM, em todos os níveis e em diversos aspectos. Entre estes aspectos destaca-se a proteção e a preservação dos dados com o uso de *ferramentas* apropriadas; a clarificação de papéis e *atribuições* em como atuar na atividade da RDM; o apoio aos *investigadores* nas questões legais e na relação formal com as *comunidades* envolvidas, possibilitando legitimar as práticas essenciais para a RDM a partir da articulação dos atores envolvidos.

Atribuições: refere-se à divisão de tarefas, de poder e de status da atividade da RDM, são formadas em razão das *políticas* elaboradas na universidade, mas também das especificidades da investigação. Assim, podem ter agentes globais em funções atribuídas pela universidade, pelas UI&D, bem como agentes locais, relativo as necessidades da investigação e das práticas essenciais para a RDM. Envolve necessariamente levar em consideração os outros atores articulados com planeamentos bem definidos para sua organização e para o correto funcionamento da RDM, indicando quem são os responsáveis pelo uso e pela gestão de recursos, como as *ferramentas* informacionais de apoio aos *investigadores*, pelo tratamento dos dados e pelas relações com as *comunidades envolvidas*, de modo a gerir e partilhar os dados de investigação com atribuições articuladas.

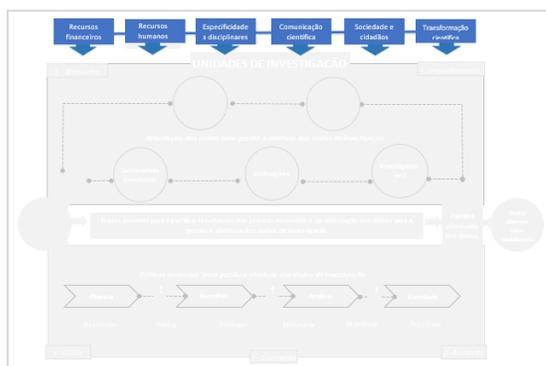
Investigador(es): são os elementos centrais para a RDM, é a partir dos investigadores que são considerados e articulados os aspectos para a gestão, partilha e reutilização dos dados, incluindo os recursos: humanos, financeiros e materiais, como as *ferramentas* e os serviços, a concepção e uso das *políticas*, as interações com as *comunidades* da investigação, bem como a elaboração da divisão do trabalho e as respectivas *atribuições*, além de realizar as práticas essenciais para a RDM e o próprio processo de investigação. Este ator, pode funcionar como elemento de ligação entre as práticas essenciais, de modo local, e os demais atores do modelo, de modo institucional, direcionado pela gestão e partilha adequada dos dados de investigação. Assim, a articulação dos atores e as práticas essenciais, dependem de modo importante das ações dos investigadores, para gerir e partilhar de forma adequada os dados de investigação.

Comunidades envolvidas: são as organizações de pessoas e de partes interessadas diretamente nos dados, na RDM e na partilha dos dados. Podem ser agentes internos ou externos à universidade que ofertam ou consomem os dados. Assim, são, ao mesmo tempo, fornecedoras e usuárias da RDM que assumem diferentes papéis em ações relativas aos dados de investigação. Podem fazer parte dessa comunidade os gestores em vários níveis da universidade, a própria comunidade científica, os grupos sociais ou indivíduos pelos quais os dados são coletados, os profissionais que desempenham ou desenvolvem ações para uso dos dados na gestão e para a partilha. Portanto, são grupos ligados à ciência ou à sociedade, que interagem com os *investigadores*, de modo a viabilizar a investigação no trabalho com os dados, através das *ferramentas*, recursos, *atribuições* de responsabilidades, *políticas* e comunicação, também fornecendo ou requerendo os dados das investigações. Sobretudo, são comunidades que compreendem os indivíduos e subgrupos que participam diretamente da investigação, do cuidado com os dados e a sua gestão, para uma partilha adequada.

7.3.7. AS FORÇAS TRANSVERSAIS INFLUENCIADORAS

Figura 48

As forças influenciadoras no modelo



Esta perspectiva é constituída por um conjunto de elementos que podem influenciar de forma indireta na atividade para a RDM das UI&D, mas de modo específico nas decisões dos investigadores sobre a gestão e a partilha dos dados. Ou seja, são forças que representam agentes ou mesmo tendências que impactam a

articulação dos atores e as práticas essenciais dos investigadores previstas no modelo, podendo gerar tensões, conflitos, bem como colaborações, que repercutem para a RDM e no ato da partilha. A transversalidade se dá no sentido destas forças não serem diretamente um dos atores articulados, mas que exercem importante influência sobre a RDM e a partilha dos dados. As seis forças transversais destacadas no modelo (retângulo com setas na cor azul, atravessado horizontalmente pela barra na mesma cor), embora não limitada a estas (o pontilhado ao final da barra representa a não limitação), são descritas a seguir.

Os **recursos econômicos** referem-se principalmente à parte financeira que influencia na realização da investigação, na gestão e na partilha dos dados, inclui os recursos humanos, financeiros, materiais e imateriais. Oriundos de matrizes variadas, como órgãos internacionais, nacionais, locais, empresas, municípios e a própria universidade. Os recursos econômicos permitem avançar ou restringir as condições para a RDM e a partilha dos dados, de modo a interferir na articulação e nas práticas a serem realizadas. Em destaque estão os recursos financeiros, que por serem limitados, finitos e dependerem de variáveis imprevistas para a sua existência, influenciam de modo importante para a atividade, a sua falta, ou mesmo o excesso sem equidade, podem causar problemas pessoais e institucionais para a RDM em um ambiente articulado e sociotécnico como as UI&D da universidade.

As **agências de fomento**, em ligação direta com a questão financeira, ganham a condição de poder influenciador através das exigências e orientações para o estabelecimento de práticas na gestão e na partilha dos dados, estimulando ou requisitando dos investigadores o compromisso para a disponibilização dos resultados da investigação em acesso aberto, de modo a gerar reações que podem repercutir na atividade da gestão e partilha. É possível que as agências de fomento sejam os elementos que mais poder tenham de influenciar o funcionamento de uma RDM, de modo institucional e pessoal. Neste sentido são muitas referências que indicam as suas políticas como influenciadoras, de modo a interferir na realização das práticas e articulações para a RDM sob o ambiente das UI&D.

As **interferências individuais** podem ser percebidas em dois níveis. Por um lado estão as influências decorrentes das percepções dos investigadores sobre o ambiente para a RDM e sobre a partilha dos dados em acesso aberto, interferindo em suas atitudes nesta última questão. Por outro lado, as influências dos gestores institucionais da universidade que possibilitam estimular ou desestimular a gestão e a partilha no ambiente. São fatores intervenientes que podem ser positivos, mas também potenciais geradores de tensões e conflitos, que, embora difíceis de mensurar e explicitar, repercutem de modo importante em como os investigadores conduzem e realizam as suas práticas de investigação, interferindo na articulação dos atores de modo institucional através de ações, ou mesmo na sua falta, para gerir e partilhar os dados adequadamente.

A **comunicação científica**, como está no cerne colaborativo da ciência, gera uma exigência institucional e pessoal no ambiente da investigação, de forma que a divulgação e a publicação dos resultados se torna parte natural do processo investigativo. Como os investigadores tendem a adaptarem-se para cumprir as exigências oriundas das atividades nesta área, a comunicação da ciência é fonte de influência e possível geradora de tensões e conflitos pessoais e institucionais, principalmente quando são combinados as exigências naturais da comunicação científica, com os interesses dos atores que a constitui, em intercâmbio com as práticas necessárias para a gestão e partilha adequada dos dados.

Uma **sociedade** que se torna cada vez mais aberta, requer transparência e mudança no tratamento da investigação financiada com recursos públicos, também em relação aos dados e ao valor que estes podem proporcionar. No mesmo sentido, a ciência caminha em direção ao engajamento dos **cidadãos** com as investigações, sendo tal colaboração aberta eficiente para a criação do conhecimento e para o processo de descoberta. Portanto, a colaboração de uma ciência cidadã e a interferência de grupos sociais ou científicos para que os dados sejam partilhados com toda a sociedade, influenciam e repercutem nas práticas essenciais e na articulação dos atores das UI&D, de modo que deve ser realizada uma gestão e disponibilização dos dados para a sociedade de forma adequada e segura.

As **inovações científicas e sociais** abrem a possibilidade para a atenção do que está acontecendo, mas também do que pode acontecer, em termos da abertura e dos recursos informacionais oriundos da ciência e da sociedade. Referem-se as influências, evoluções e criações de novas tecnologias, ferramentas e inovações metodológicas que, compatibilizadas com a filosofia aberta, podem ser úteis ou necessárias para o avanço eficiente da gestão e partilha dos dados. Estas inovações, quando implementadas, são potenciais fontes de tensões e conflitos institucionais e individuais, pois no confronto com o modo usual ou já estabilizado da realização das práticas da RDM para a partilha dos dados, podem influenciar com uma desestabilização e interferência que deve ser levada em consideração para que a atividade possa se adaptar e readequar a tais inovações, quando assim forem necessárias.

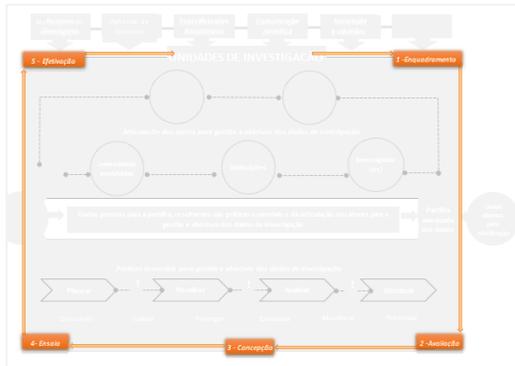
As forças influenciadoras transversais têm a sua função no modelo, orientadas pela interferência na RDM e na partilha dos dados. Cabe, contudo, ressaltar a sua não limitação e a mutabilidade que as forças podem sofrer com o tempo, bem como as eventuais transformações

da ciência e sociedade. Portanto, a percepção desta perspectiva é direcionada no modelo para a RDM, para ser realizada e organizada pela metodologia para a adaptação contextual.

7.3.8. A METODOLOGIA PARA ADAPTAÇÃO CONTEXTUAL

Figura 49

A adaptação contextual da RDM no modelo



A perspectiva da metodologia para a adaptação contextual da atividade da RDM ao ambiente, permite que as práticas essenciais e os atores articulados sejam observados em conjunto, bem como todas as outras perspectivas do modelo, de modo a possibilitar o ajuste das especificidades das investigações e da heterogeneidade dos

dados para aplicar a RDM proposta, considerando o contexto das UI&D e da universidade.

Relacionada com as perspectivas do modelo, a metodologia permite implementar a estrutura de modo que o global e o local se relacionem. Ou seja, a adaptação contextual da RDM ao ambiente acontece nos dois sentidos abordados no modelo: 1) no sentido horizontal e local, onde a adaptação da RDM é realizada observando as perspectivas do modelo em relação as especificidades do ambiente da investigação; e, 2) no sentido vertical e global, onde a adaptação acontece tendo por consideração o ambiente institucional da universidade. Assim, a conexão entre o global e o local pode ser realizada para o apoio da universidade ao ambiente da investigação e o ambiente da investigação pode ser adaptado para suprir as necessidades informacionais da universidade. Para tanto, o modelo sugere o uso de um método de intervenção formativa, com ações sucessivas (caixas retangulares e setas na cor laranja), organizadas e direcionadas para esta adaptação contextual.

A intervenção sugerida parte do pressuposto que em todos os ambientes de investigação existe uma forma de RDM, mas que precisa ser realizada sob as características de um novo modelo, orientado para a articulação dos atores, para a realização de práticas essenciais e para uma partilha adequada dos dados. Assim, a intervenção conduz de modo organizado à adaptação contextual do modelo para a RDM ao ambiente, realizada através

de um processo de diálogo e colaboração multivocal, onde participantes-chave tomam iniciativa e são protagonistas para a adaptação da própria atividade, através de um ponto de partida, seguido de mais cinco etapas, conforme segue as explicações.

Identificação das perspectivas do modelo – é o ponto de partida da metodologia, serve para reconhecer e visualizar os atores da RDM no ambiente, bem como as práticas essenciais, as forças transversais, as características dos dados e do próprio ambiente. Esta identificação possibilita a análise das relações, influências e práticas utilizadas para a gestão e partilha dos dados. Em uma análise inicial, ao identificar e arranjar os atores existentes de forma articulada, obtém-se a visão da RDM atual, para posterior enquadramento dos problemas e proposições de soluções para que a RDM seja adaptada ao ambiente.

1) Enquadramento – a partir da visão atual da RDM, um quadro da articulação dos atores, das práticas essenciais, das forças transversais e das características dos dados e do ambiente é desenvolvido. O enquadramento contém os aspectos importantes a serem observados e as dificuldades adaptativas atuais no caminho para a nova RDM adaptada ao ambiente. Tal quadro produzido observa os fatores positivos, mas o foco principal são as situações problemáticas verificadas no contexto local e global para as articulações e práticas em uma partilha adequada dos dados. Verificado o contexto, associado à percepção do estado atual da RDM no ambiente, gera-se um quadro analisável dos problemas para um processo reflexivo, de modo a realizar a adaptação contextual da RDM.

2) Avaliação – esta etapa caracteriza-se por ser um diagnóstico aprofundado dos problemas direcionados à RDM atual, ou seja, a partir do quadro gerado na fase anterior, realiza-se a análise para concepção da RDM adaptada ao ambiente. Esta avaliação indica a superação dos problemas identificados, possibilitando a projeção da RDM objetivada. Assim, através dos exames acerca das dificuldades, produz-se uma imagem detalhada em relação às causas das adversidades encontradas, bem como do papel das perspectivas do modelo em relação aos problemas. Propõe-se então, um modo para superação de tais problemas, equacionando as dificuldades encontradas e criando soluções inovadoras que possibilitam a contextualização da RDM para que as articulações e as práticas, sejam adaptadas ao ambiente e os dados sejam partilhados adequadamente. Assim as soluções propostas são os elementos necessários para conceber a RDM adaptada.

3) Concepção – é a modelagem da atividade da RDM para adaptação ao ambiente. Através desta modelagem é possível conceber uma nova estrutura para a atividade, na expectativa da solução dos problemas observados na etapa de avaliação, possibilitando uma nova geração de ideias para adaptação da RDM. Ou seja, a concepção de uma RDM adaptada ao ambiente é estabelecida nas ideias sobre uma RDM futura, construindo um modelo explícito e simplificado. Esta concepção objetiva mostrar a situação problemática e oferecer uma perspectiva prática para resolvê-la, influenciando a atividade para sua adaptação. São aqui definidas novas formas e uma nova solução é materializada, mediante a produção de versões e modificações das ideias originais, até que se chegue a um novo modelo aceitável. Tal modelo também deve permitir que a nova forma de RDM idealizada, possa ser legitimada através de ensaios sucessivos para posterior efetivação.

4) Ensaios – nesta fase, o objetivo é levar as ideias abstratas levantadas e modeladas na etapa de concepção para um nível mais concreto e mais próximo de ser realizado, podendo ser testadas e/ou validadas na atividade real através de um programa piloto, que precede a efetivação da nova RDM. Os testes do novo modelo ocorrem com o objetivo da exequibilidade da relação entre a nova RDM adaptada e a RDM atual no ambiente, para antecipar o que pode ser auxiliado na adaptação, bem como identificar os obstáculos que se podem encontrar na atividade real. Parte da tarefa dos ensaios é continuar a concepção do modelo ao nível dos detalhes. Tendo em vista que durante o ensaio novos problemas podem surgir, avalia-se se o fluxo segue para a etapa de efetivação ou retorna para o nível da concepção para eventuais correções, novo ensaio e posterior efetivação.

5) Efetivação – são as ações para pôr em produção o modelo para a RDM no ambiente. Ou seja, é a efetivação da articulação dos atores e das práticas essenciais, levando-se em consideração todo o contexto, através das características dos dados, das forças transversais e das especificidades investigativas do ambiente para uma partilha adequada. A efetivação põe em prática o modelo através de um processo gradativo e adaptativo, onde os resultados também devem ser avaliados em uma situação da realidade ampliada, não mais em um piloto. A expectativa em pôr o modelo de forma alargada no ambiente é de que se realize um processo de acomodação. Assim, enquanto as perspectivas da RDM são utilizadas, avalia-se a necessidade de alterações para ajuste à nova realidade até que se torne estável, pois ainda

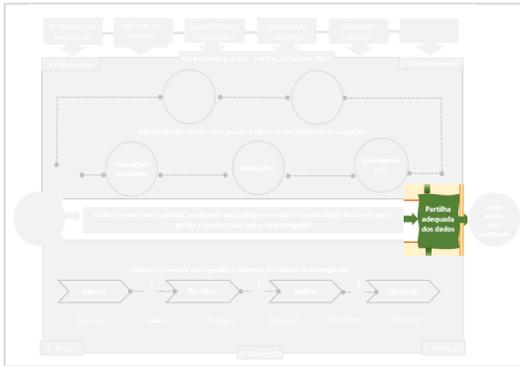
pode ocorrer problemas de adaptação entre a antiga RDM e a nova forma de fazer e pensar a gestão dos dados. Efetivada, a RDM adapta-se ao ambiente de modo que o contexto local e global estejam influenciados de maneira recíproca e colaborativa sob o mesmo objeto.

Com a RDM estável e adaptada, alcança-se a forma ideal para que todas as perspectivas sejam consideradas em conjunto, de maneira organizada e de modo que esta conjunção proporcione uma gestão necessária para a realização de uma partilha adequada dos dados de investigação.

7.3.9. A PARTILHA ADEQUADA DOS DADOS DE INVESTIGAÇÃO

Figura 50

Uma partilha adequada dos dados de investigação.



A perspectiva para uma partilha adequada dos dados de investigação (identificada no modelo pela caixa retangular na cor verde), é visualizada sob o grau de inserção e a centralidade que esta partilha possui para a RDM. Assim, a importância da “partilha adequada” pode ser vista de duas maneiras no modelo para a RDM.

A primeira maneira é como base orientadora para a RDM e ao mesmo tempo receptora dos dados. Como base orientadora para a RDM, influencia as perspectivas da articulação dos atores e das práticas essenciais de gestão, para que a partilha dos dados possa ser assegurada pelas infraestruturas digitais e sempre que possível na forma do acesso aberto, com todo o processo objetivado para a reutilização dos dados. Nesse sentido, a articulação dos atores e as práticas essenciais, se conectam à “partilha adequada” de modo a serem assim fundamentadas e orientadas (representada através da barra de ligação na cor gradiente ligado ao retângulo). Esta conexão significa que a “partilha adequada” é a base para realização das duas perspectivas. Como base receptora, uma partilha adequada é o alvo a ser atingido pelos dados (setas direcionais na cor verde) oriundos da articulação dos atores e das práticas essenciais, tendo como finalidade serem transformados em dados abertos e consequentemente estejam disponíveis para serem reutilizados.

A segunda maneira para visualizar a partilha adequada no modelo é como elo de comunicação entre as práticas essenciais e a articulação dos atores (representado pelas duas barras de ligação), que, por intermédio de interlocutores ou do investigador, pode conectar o global e o local, com a participação na investigação e na interação com os atores institucionais.

Portanto, seja como meio de ligação ou como base para a fundamentação e orientação, ou ainda na influência sobre as outras perspectivas especificadas, a partilha adequada dos dados de investigação é o cerne do modelo e o objeto da atividade para a RDM ser desenvolvida através desta estrutura modelar no ambiente da universidade.

Apresentado neste capítulo de forma gráfica e textual, o modelo para a RDM proposto, derivado das abordagens teórica e empírica que foram realizadas gradativamente ao longo deste estudo, deve ser definida como flexível, pois direcionado pela metodologia para a adaptação contextual, tem na sua construção multivocal a possibilidade de ajustes necessários em todas as sete perspectivas formadoras do modelo. Em essência, o modelo pretende ser aplicado de modo direcionado através de padrões estabelecidos, mas também para permitir acrescentar métodos e soluções pertinentes na resolução dos problemas apresentados ou potencialidades a serem desenvolvidas.

É importante destacar que, conforme planeado nesta investigação, o modelo para a RDM foi desenhado para a implementação nas UI&D da universidade, mas em uma perspectiva alargada e tendo em vista os recursos para a adaptabilidade, é possível pensar que este modelo possa ser aplicado tanto em um projeto de investigação, quanto ampliado para uma rede de atividades para RDM, cuja ampliação pode ser gradativamente expandida, de forma organizada e de modo a abranger todo ambiente da universidade.

Por fim, como o modelo não pretende prever todas as possibilidades para uma RDM no objetivo proposto, as questões abertas e que sejam encontradas quando da sua utilização, devem ser resolvidas à medida em que o modelo é aplicado, orientadas pela adaptação contextual e de modo que em todos os casos na sua aplicação resulte na finalidade prevista, qual seja: uma RDM realizada através das articulações dos atores e das práticas essenciais para a gestão e partilha adequada dos dados no ambiente da universidade.

CONCLUSÕES

O cumprimento do objetivo geral do presente estudo resultou na construção de um modelo para a RDM (*Research Data Management*) no ambiente da universidade, que possibilita uma partilha adequada dos dados, com base na articulação dos seus atores e no uso de práticas essenciais. Para tanto, o caminho percorrido foi direcionado pelos objetivos específicos deste trabalho, que proporcionaram de modo teórico a contextualização do ambiente da RDM na universidade, para em seguida qualificar uma partilha adequada dos dados de investigação e proceder à identificação das práticas essenciais para a RDM, e ainda viabilizar uma proposta para articular os atores; adaptar a RDM ao contexto e conceber o nexo teórico do estudo. Por fim, de modo empírico, foi caracterizado o ambiente das Unidades de Investigação e Desenvolvimento (UI&D) na Universidade de Coimbra (UC), também os atores e as práticas envolvidas, bem como os hábitos para a partilha dos dados em acesso aberto.

O esforço para obter o resultado desta investigação, teve como origem o questionamento sobre o modo na qual uma RDM poderia ser estruturada para proporcionar uma partilha adequada dos dados no ambiente da universidade, tendo por base o uso de atores articulados e de práticas consideradas como essenciais. Para responder à questão levantada e para o cumprimento do objetivo geral da tese, o trabalho foi fundamentado pela ciência aberta como base teórica, juntamente com o estudo da teoria da atividade e de modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação, que associados e confrontados com o campo empírico através dos dados coletados nas UI&D da UC, resultaram no modelo apresentado.

O modelo para a RDM proposto conta com sete perspectivas que, juntas, cumprem o objetivo geral desta tese, quais sejam tais perspectivas: 1) as unidades de investigação; 2) os dados a serem geridos e abertos para a reutilização; 3) a partilha adequada dos dados de investigação; 4) as práticas essenciais para a gestão e a partilha dos dados de investigação; 5) a articulação dos atores para a gerir e partilhar os dados de investigação; 6) as forças transversais influenciadoras; e, 7) a metodologia para adaptação contextual.

Portanto, a própria estrutura do modelo apresentado é o modo indicado por este estudo para responder à questão principal, percebida tanto através da observação de tal estrutura modelar, quanto pela sua explicação textual. As conclusões decorrentes de cada um dos objetivos específicos alcançados também ratificam a resposta da questão de modo detalhado, além de confirmar a hipótese defendida no início desta investigação, de que uma RDM estará mais próxima de uma partilha adequada dos dados no contexto da universidade, se estiver organizada segundo um modelo que vincule os atores no ambiente, associado ao uso de práticas essenciais para a gestão dos dados. É nesse sentido que as conclusões dos objetivos específicos e do objetivo geral alcançados são apresentadas a seguir.

Objetivos alcançados

1. Contextualização do ambiente temático da RDM na universidade: o alcance deste primeiro objetivo proporcionou o apoio necessário para a fundamentação inicial e desenvolvimento do trabalho, de forma direta corroborou para a resposta da questão levantada e na formação do modelo proposto, ao exibir um panorama teórico e relacional entre o ambiente da RDM e a universidade, mas também em relação a este ambiente e a identificação dos atores e as características dos dados de investigação, possibilitando assim constatar:

- A necessidade da RDM na universidade;
- As demandas de ações para a coordenação e articulação entre os atores da RDM;
- Os desafios técnicos e sociológicos a serem observados no ambiente;
- A abrangência da RDM na universidade através de práticas globais e locais;
- A utilidade dos dados, quando caracterizados a partir da RDM na universidade;
- Os desafios para gerir dados heterogêneos, fragmentados e “invisíveis”;
- A complexidade e a diversidade do ambiente; dos atores e dos dados; e
- A necessidade da observação das especificidades do ambiente dos dados e dos atores.

Os desafios e as necessidades constatadas em relação à gestão, aos atores, aos dados de investigação e ao ambiente são supridos no modelo apresentado, pois este contém as práticas necessárias para que uma RDM na universidade atente aos desafios técnicos e sociológicos, também permitindo organizar e articular os atores de modo local e global e possibilitando uma solução adaptativa para os problemas causados pela complexidade e

heterogeneidade, sob as características dos dados e dos atores no ambiente. A partir do que foi constatado, pode-se perceber que a problemática encontrada nesta contextualização se coaduna com a maior parte dos problemas levantados na tese, sendo fonte de direcionamento para soluções a partir deste ambiente temático.

O resultado do panorama relacional entre o ambiente temático e as características dos dados de investigação foi essencial para o modelo proposto, no sentido de que, além de contribuir para ratificar que os dados possuem características diversas, heterogêneas e que repercutem para a sua fragmentação e invisibilidade no contexto da universidade, trouxe para o modelo a necessidade da observação destes dados de forma específica e contextual, bem como a importância da identificação das suas características para a realização da RDM. No modelo proposto, esta observação, identificação e caracterização específica e ambiental é realizada através da análise orientada pela perspectiva da metodologia contextual.

Na relação entre o ambiente temático e os atores, inicialmente foram identificados e sugeridos que os atores fossem associados e agrupados através das classes: pessoas, políticas e infraestruturas. No entanto, estas classes serviram para indicar um modo possível de agrupamento, pois no modelo para a RDM proposto, estes atores ganharam novas nomenclaturas e tiveram suas funções ampliadas na perspectiva da articulação com as práticas essenciais, com as forças influenciadoras e na adaptação da RDM ao ambiente contextual.

O primeiro objetivo específico, também respondeu parte da questão da tese, ao sugerir que os atores possam ser identificados por classes e apontar a necessidade de agrupá-los sob um contexto específico a ser explorado, sem desconsiderar a possibilidade da variação entre áreas disciplinares diferentes. Assim, no modelo proposto, pôde-se constituir classes genéricas de atores, ao mesmo tempo que se leva em consideração as características dos dados de investigação para o ambiente complexo e heterogêneo da universidade, atentando para as especificidades das áreas investigativas juntamente com o contexto explorado.

É possível concluir que a contextualização do ambiente temático da RDM na universidade, contribuiu de forma efetiva para a resposta da questão principal em relação à identificação dos atores, bem como para a construção do modelo, no sentido de influenciar que tal modelo proporcione a flexibilidade e a atenção com as especificidades dos atores,

com os dados de investigação, com o ambiente, e no relacionamento entre eles, mas também a possibilidade de olhar em dois sentidos, ou seja, no global, representado no modelo pela instituição e pelos atores articulados, bem como para o local, representado pelas práticas essenciais, ambos direcionados para a gestão e para a partilha dos dados de investigação.

2. Caracterizar uma partilha adequada dos dados: a consecução do segundo objetivo, decorre da organização relacional e do agrupamento em aspectos temáticos dos princípios e práticas da ciência aberta, no contexto do acesso aberto e da abertura dos dados, observados também as tensões e conflitos consequentes, resultando nos elementos encontrados para a realização de uma partilha adequada dos dados de investigação.

Portanto, a construção acima possibilitou reunir as informações necessárias para caracterizar o que seja uma partilha adequada dos dados na universidade, que foi definida como: um processo sistêmico e organizado para a abertura dos dados de investigação, realizada por um conjunto de atores associados, centrada no investigador como ator principal, sob o qual devem ser observadas as influências cognitivas, contextuais e sociotécnicas. Fundamentada no uso relacional e intencional dos princípios e práticas da ciência aberta, o processo para a partilha dos dados deve ser suficientemente flexível para abarcar as especificidades das áreas de conhecimento, as orientações e as exigências das políticas, bem como as documentações que facilitam o uso atual e futuro dos dados, sendo assegurada por infraestruturas digitais e sempre que possível na forma do acesso aberto, com todo o processo objetivado para a reutilização dos dados de investigação. Esta definição também responde à questão complementar: o que é uma partilha adequada?

As informações necessárias que permitiram o alcance deste objetivo específico, apontam e sustentam os seguintes elementos presentes no modelo para a RDM:

- O acesso aberto para a partilha dos dados já processados na investigação;
- As diversas influências que podem afetar o investigador;
- As especificidades e os agentes relacionais do contexto;
- Os repositórios ou infraestruturas digitais adequadas para a partilha;
- A reutilização dos dados pelos investigadores e pela sociedade;
- A centralidade e a atenção primária para o investigador na investigação local;

- A ação da partilha como sendo parte de um processo maior;
- As formas e os meios considerados ideais para a partilha;
- Os princípios e as práticas da ciência aberta;
- A organização relacional dos princípios e práticas da ciência aberta;
- O movimento para a partilha como resultado dos princípios e práticas da ciência aberta;
- As tensões e conflitos gerados no ambiente;
- As barreiras levantadas para a partilha dos dados;
- Os benefícios que justificam a partilha dos dados;
- As especificidades do contexto;
- As diferenças disciplinares do ambiente;
- Os agentes do sistema de comunicação científica;
- Os atores necessários para abertura dos dados;
- A relevância dos aspectos técnicos e sociais;
- Uma organização necessária de forma lógica e proposital;
- A atenção a cognição e ao contexto do investigador;
- A flexibilidade para adaptação as peculiaridades do ambiente;
- A sistematização dos atores; e
- O direcionamento para a reutilização dos dados.

O modelo para a RDM contém e atende os elementos e as considerações resultantes que caracterizam uma partilha adequada dos dados, permitindo assim efetivar um processo de sistematização organizado para a sua gestão – resultado do uso intencional e relacional dos princípios e práticas da ciência aberta –, materializado por intermédio das perspectivas da articulação dos atores e da utilização das práticas essenciais para a RDM.

As considerações resultantes são viabilizadas no modelo, principalmente na perspectiva da metodologia para a adaptação contextual, de modo que a associação dos atores seja centrada no investigador como ator principal, considerando que o ponto de vista inicial para a utilização da metodologia é realizado a partir do investigador. A metodologia, além de possibilitar o processo de organização, também propõe que a RDM seja adaptada ao ambiente, permitindo uma flexibilidade suficiente para que o movimento para a partilha

dos dados possa abarcar as especificidades das áreas de conhecimento e considere as influências sociotécnicas, cognitivas e contextuais sobre o investigador, onde tais influências podem gerar tensões e conflitos, interferindo assim no ato da partilha dos dados.

Está também previsto no modelo, o cuidado com outras influências do contexto, como, por exemplo, as exigências das agências de fomento e da comunicação da ciência, representadas na perspectiva das forças transversais, que, de modo importante, são observadas para a realização da partilha. No mesmo sentido, é verificada, através da perspectiva das práticas essenciais, a atenção direcionada ao processo, e não apenas para o ato da partilha. A atenção ao processo, propicia a gestão e a disponibilização dos dados através das infraestruturas digitais e quando possível na forma do acesso aberto, e ainda, a possibilidade de reutilização dos dados, viabilizadas de modo específico pelas práticas essenciais nomeadas de descrição e monitoramento, facilitando o uso atual e futuro dos dados de investigação, bem como os benefícios previstos para a ciência e para a sociedade.

Como consequência da realização deste objetivo específico, os benefícios proporcionados pela partilha adequada dos dados de investigação, a serem garantidos através do modelo, podem ser vistos na exequibilidade da transparência, das inovações e descobertas, permitindo também que a ciência avance e que a reutilização dos dados colabore com a eficiência científica e com uma comunicação adaptada a novos paradigmas, podendo gerar a interdisciplinaridade e a igualdade de oportunidades para os investigadores, assim como a redução da dualidade nas investigações e uma melhor explicação da ciência.

Outros benefícios oriundos do cumprimento deste objetivo, que estão sob os princípios e práticas da ciência aberta, podem ser percebidos através do modelo de modo determinante como, por exemplo, o aspecto colaborativo, visto na forma participativa e inclusiva que permite a articulação entre os investigadores, entre os investigadores e a comunidade acadêmica e entre ambos e a sociedade. O modelo também repercute nos aspectos pessoal e social, através do princípio da disponibilidade e do acesso para todos.

No aspecto social, uma partilha adequada dos dados pode ser revertida no bem-estar para a sociedade e em soluções para o interesse coletivo, também na redução de desigualdades informacionais e na possibilidade de soluções para os problemas sociais. O

aspecto pessoal é percebido na perspectiva da articulação dos atores, na definição de responsabilidades e nos apoios necessários para que o investigador possa partilhar os dados de forma adequada. Aspectos importantes que remetem à legalidade e à sustentabilidade, bem como a parte técnica e material, são atendidos através da perspectiva das práticas essenciais, quando é proporcionado que os dados sejam descritos, protegidos, monitorados, estruturados e preservados de modo interrelacional com os demais elementos que sustentam a RDM.

É possível concluir, que a realização de uma partilha adequada dos dados de investigação através do modelo para a RDM, permite cumprir vários aspectos dos princípios e práticas da ciência aberta, proporcionando os benefícios decorrentes, porém, sem desconsiderar as características do contexto e as dificuldades para a realização da partilha no ambiente.

3. Identificar as práticas essenciais para a gestão dos dados: foi cumprido como terceiro objetivo específico, sucedendo também na resposta da questão complementar que deseja saber quais são tais práticas essenciais. O estudo de nove modelos de ciclo de vida resultou na identificação, seleção e descrição de etapas sequenciais e de etapas contextuais, que juntas e associadas, formaram as práticas essenciais a serem utilizadas para gerir e partilhar adequadamente os dados de investigação no ambiente da universidade.

Este objetivo alcançado caracterizou as práticas essenciais como: um conjunto de ações para gerir e partilhar os dados de investigação de forma adequada, sendo constituído por um fluxo de etapas sequenciais iniciadas com o planeamento e realizadas através da recolha e análise dos dados, tendo a sua finalização com o ato da partilha. O fluxo é permeado por práticas contextuais que abrangem duas ou mais etapas sequenciais, de modo a garantir a descrição e a monitoração dos dados, mas também a constituir a proteção legal e técnica necessária, bem como os recursos estruturais para preservação, gestão e a partilha dos dados de investigação. Sendo direcionadas ao acesso aberto, sob os princípios da ciência aberta e objetivada para a reutilização dos dados, as práticas essenciais devem ser associadas ao ambiente e ao conhecimento contextual, proporcionando também a distinção entre as práticas da RDM e da curadoria digital, de modo a evitar a justaposição de trabalhos.

Este terceiro objetivo atingido permitiu que o modelo para a RDM acolhesse a definição apresentada, de modo que as práticas essenciais fossem representadas por etapas, tendo o processo de investigação como seu eixo central e viabilizado por um fluxo de práticas sequenciais no que se refere ao planeamento, recolha, análise e partilha dos dados. Possibilitou ainda, que a curadoria dos dados estivesse presente, promovida pelas práticas contextuais para descrever, proteger, estruturar, monitorar e preservar não somente os dados, mas também os materiais e demais elementos necessários para realizar uma RDM direcionada ao acesso aberto e aos dados abertos, sob princípios e práticas da ciência aberta.

Além do destaque no modelo para o fluxo das práticas sequenciais que remetem ao processo de investigação, são evidenciadas as etapas contextuais realizadas durante este processo, orientadas principalmente à curadoria dos dados. As práticas contextuais possibilitam que a RDM seja monitorada, através das relações institucionais, com a devida atenção para as estruturas e de modo que as fases do processo de investigação, juntamente com os itens para a curadoria, como a descrição, proteção, preservação dos dados e dos materiais subjacentes, sejam garantidos. Assim, o conjunto de ações sob estes agrupamentos de etapas constituem-se nas práticas essenciais para que a RDM resulte em uma partilha adequada dos dados, com ações a partir do investigador e da investigação, mas também através de uma conexão necessária com os demais atores articulatórios.

A conexão articulatória dos atores com as práticas essenciais também é cumprida no modelo através da perspectiva da metodologia para adaptação contextual, uma vez que a proposta da metodologia permite o ajuste das especificidades do ambiente, possibilitando que as práticas essenciais e os atores articulados sejam observados e utilizados em conjunto. A observação de forma associada é importante, pois, vistos de forma isolada, tais práticas não são suficientes para a obtenção da RDM objetivada, já que é necessário dar relevo ao ambiente da universidade, com os seus atores diversos, e ao contexto específico da investigação. Assim, conforme é proposto nas perspectivas do modelo, a utilização sistêmica ampla dos atores articulados, observadas as características do ambiente, bem como as influências para realização da gestão e da partilha adequada dos dados, são necessárias para maximizar o uso das práticas essenciais identificadas.

Portanto, conclui-se que o alcance deste objetivo específico, permite que as práticas essenciais para a RDM a serem utilizadas através do modelo, compreendam os itens necessários para uma gestão e curadoria pertinente no ambiente, a partir da investigação e do investigador, tendo por consideração os atores no contexto da universidade. Assim, a aplicação de tais práticas essenciais, de forma sistêmica, juntamente com os atores articulados e com a observação das especificidades da investigação, resultam na potencialização da RDM na universidade e em uma partilha adequada dos dados.

4. Estabelecer uma forma de articular os atores para gerir os dados: proporcionou que o modelo para a RDM pudesse identificar, arranjar e articular de forma sistêmica os atores e possibilitou a adaptação da atividade da RDM ao ambiente das UI&D, podendo ser extensivo a outros contextos de investigação da universidade.

Em resposta às questões complementares que buscaram saber como deve ser a articulação dos atores para a RDM e como a gestão dos dados pode adaptar-se ao ambiente complexo e heterogêneo da universidade, este objetivo alcançado definiu a articulação dos atores como: um tipo de articulação sistêmica que possibilita a realização de práticas essenciais para a RDM e uma partilha adequada dos dados, através do relacionamento entre os elementos humanos e não humanos, constituídos a partir dos componentes envolvidos para a RDM na universidade, sob o acesso aberto e fundamentado na ciência aberta. Para tanto, os atores são agrupados por classes de investigadores, comunidades, ferramentas, políticas e atribuições para o trabalho, viabilizados por um método interventivo de aprendizagem que permite a adaptação ao ambiente, mediante um processo onde os atores são sistematizados, os problemas enquadrados e avaliados para conceber soluções, que após modeladas e postas em testes podem ser efetivadas para o uso, ou retornar ao processo para aperfeiçoamento. Quando efetivado, o processo resulta em atores articulados e uma RDM adaptada ao ambiente, para a realização de práticas essenciais de gestão e uma partilha adequada dos dados de investigação.

O estabelecimento da articulação dos atores está inserido no modelo final para a RDM, novamente através da perspectiva da metodologia para a adaptação contextual. A utilização da metodologia realiza o arranjo e a articulação sistêmica dos atores, mas também põe em análise as outras perspectivas, como a verificação das práticas essenciais para a RDM, a influência realizada sobre a partilha dos dados e a caracterização dos próprios dados

e do ambiente, ajudando a minimizar os problemas da complexidade e da heterogeneidade na RDM na universidade. Isto pode acontecer porque a forma para a articulação dos atores que o modelo propõe, é aplicada através de uma construção interventiva e multivocal que sistematiza os atores e a estrutura da RDM, de modo que as especificidades contextuais sejam adaptadas e onde o local e o global se encontram através desta construção, possibilitando que se apoiem e influenciem reciprocamente tornando os dados conhecidos.

No sentido exposto, são realizadas ações através de etapas que proporcionam a identificação e o arranjo articulado dos atores, juntamente com a análise e adaptação da RDM ao contexto. Assim, de modo propositivo, concebem-se soluções para a RDM que, efetivada dentro do processo estabelecido, pode resultar em uma gestão dos dados modificada e adaptada ao contexto complexo e heterogêneo da investigação na universidade. Portanto, este objetivo específico garante uma forma de articular os atores para a gestão e a partilha adequada dos dados, proporcionando uma atenção específica aos dados de investigação, bem como as forças que influenciam a partilha e a utilização de práticas essenciais, adaptando a RDM ao ambiente da investigação.

Conclui-se que este objetivo alcançado e a forma de articulação encontrada, permite que o modelo para a RDM não somente articule os atores para a gestão e a partilha dos dados, mas possibilite que o local e o global se relacionem, de modo que o contexto institucional tenha o conhecimento necessário sobre o ambiente local da investigação e sobre os dados geridos. Em sentido inverso, o ambiente local da investigação pode ter o suporte institucional necessário para realizar a RDM de forma pertinente e apoiada pela universidade. Portanto, este objetivo viabiliza que o modelo possa atender aos problemas e às especificidades para a RDM no ambiente multidisciplinar e heterogêneo da universidade, através da articulação dos atores e da adaptação da RDM ao seu contexto. Nesse sentido, podem ser parcialmente supridas as dificuldades na articulação e nas práticas isoladas para a RDM, bem como na fragmentação e na invisibilidade dos dados. Mas também, a forma de articulação dos atores aqui alcançada, potencializa todo o modelo para a RDM de modo que a partir deste objetivo, seja possível sua implementação e expansão para obter como resultado uma RDM realizada no ambiente da universidade, onde através das articulações dos atores e das práticas essenciais têm-se como produto uma partilha adequada dos dados.

5. Propor um nexu teórico entre a articulação dos atores, as práticas essenciais para a RDM e uma partilha adequada dos dados: cumpre a relação conceitual estabelecida entre estes três eixos de estudo, fundamentados respectivamente pela teoria da atividade; pelos modelos de ciclos de vida que representam os fluxos da RDM, da curadoria e dos dados de investigação; e, de forma basilar, sob a orientação da ciência aberta.

O cumprimento deste objetivo assegurou que o modelo para a RDM estivesse firmado em bases conceituais sólidas através de um nexu relacional, tal nexu tem principalmente na ciência aberta e no acesso aberto a interseção entre a partilha dos dados e as práticas para a RDM, mas como tais práticas e a partilha são realizadas por atores, estes, quando articulados também funcionam como um elemento integrador entre ambos, que os conecta, e são por eles constituídos em interseção com o acesso aberto. Assim, toda a relação tem no acesso aberto o seu nexu, orientado fundamentalmente pela ciência aberta.

Para se chegar à relação conceitual proposta, foi realizada a análise da literatura obtida durante a fundamentação teórica da tese, resultando no nexu objetivado. Além de fundar as bases do modelo para a RDM na universidade, tal análise também foi utilizada para caracterizar o que seja uma partilha “adequada” dos dados; conceituar as práticas “essenciais” para a RDM; fundamentar a articulação dos atores na forma sistêmica e a metodologia para a adaptação contextual como um caminho para a utilização prática do modelo. O alcance deste objetivo permitiu que os eixos de estudo exerçam suas funções em conjunto, com um propósito único e direcionado em uma relação de interdependência e complementaridade, sob o ambiente da universidade. Além disso, os dados coletados por meios das entrevistas foram confrontados e discutidos com este construto, de modo a proporcionar as direções que conceberam o modelo proposto.

Conclui-se que a relação entre os eixos de estudo, além de mostrar a existência de um modelo conceitual seguro e interrelacionado, sustenta as sete perspectivas constituintes do modelo para a RDM na universidade, já identificadas anteriormente e ratificadas abaixo:

- A partilha adequada dos dados de investigação;
- As práticas essenciais para a gestão e partilha adequada dos dados de investigação;
- Os atores articulados para a gestão e partilha adequada dos dados de investigação;

- Atores que influenciam como forças que interferem no propósito da partilha;
- Os dados de investigação a serem geridos e abertos para a reutilização;
- O ambiente da investigação na universidade; e
- Uma metodologia para adaptação contextual.

6. Perceber o ambiente e as práticas para a RDM no campo empírico, juntamente com os dados de investigação, os atores relacionais e a partilha dos dados:

foi o objetivo cumprido com base na compreensão dos investigadores das UI&D da UC, sendo direcionado pelos três eixos de estudo da tese - a articulação dos atores, as práticas para a RDM e a partilha dos dados de investigação, observados no ambiente. Os dados que foram coletados através de entrevistas e confrontados com a análise da literatura, ajudaram a compor o modelo para a RDM através das seguintes conclusões e recomendações:

O ambiente da RDM

- O ambiente das UI&D onde a investigação está alocada deve ser compreendido, pois as dificuldades institucionais e as diferenças entre as UI&D influenciam na RDM;
- A minimização das dificuldades e das diferenças no ambiente das UI&D são medidas necessárias para a RDM na universidade;
- A compreensão das especificidades das investigações, no ambiente diverso da universidade, é fundamental para poder gerir e partilhar os dados;
- A percepção do ambiente deve abranger as investigações e as UI&D, em uma abordagem multidimensional entre o individual e o institucional;
- O modelo resultante deve possibilitar um viés generalista do ambiente das UI&D, sem deixar de considerar as especificidades e as diferenças disciplinares das investigações; e
- O investigador é o ator principal, e deve ser observado sob as influências do ambiente que repercutem nas práticas para a gestão e partilha dos dados.

Em boa parte já ratificadas na teoria, estas conclusões sobre o ambiente da investigação estão contempladas no modelo pela perspectiva da metodologia para a adaptação contextual, sobre a qual recai a função de identificar e compreender o ambiente, seja nas especificidades da investigação, de modo geral na UI&D, ou mesmo em um contexto maior da universidade. São ações realizadas através de uma abordagem

multidimensional entre o individual e o institucional, considerando o investigador como ator primário a ser observado na apresentação de soluções para a RDM, expandindo esta observação a outros agentes envolvidos no contexto da universidade, de modo a minimizar as dificuldades e orientar as providências para enfrentar as diferenças encontradas no ambiente.

Portanto, o modelo pode ser utilizado em um viés generalista no âmbito das UI&D, sem deixar de considerar as especificidades e as diferenças disciplinares das investigações. No mesmo sentido de solução e em conjunto com a metodologia, as influências sobre o investigador podem ser identificadas e tratadas na perspectiva das forças transversais, e ainda, a caracterização do ambiente pode ser realizada na perspectiva das práticas essenciais, com as etapas de planejamento e descrição. A compreensão e a identificação do contexto e seus vários níveis de atuação e influências, permite que outras perspectivas sejam providas destas informações acerca da conjuntura diversa da investigação na universidade, de modo a permitir gerir e partilhar adequadamente os dados de investigação.

Os dados de investigação:

- Os dados gerados nas UI&D necessitam ser caracterizados, estudados e organizados para perceber o valor que possuem, como e onde armazená-los, o que fazer com eles e os danos que suas perdas podem causar, bem como evidenciar os benefícios do seu uso;
- A caracterização dos dados deve ser acompanhada por uma leitura do ambiente da investigação para que possam ser contextualizados;
- Para gerir volumes, tipos e formatos de dados heterogêneos, a base de suporte deve ser gradualmente desenvolvida, alcançando os itens comuns e mais simples, seguindo para uma maior complexidade e abrangência;
- Um ator primordial a ser considerado é o investigador, como detentor de informações essenciais, para sugerir onde os dados podem ou poderiam ser reutilizados;
- O uso ou reuso dos dados pela sociedade ou de forma interdisciplinar, precisa de informação representativa para que sejam objetos informacionais relevantes;
- Para a reutilização dos dados, é necessário que as práticas na RDM para a partilha sejam realizadas com esse objetivo, sendo preciso uma estrutura de organização e apoio, de modo que o investigador possa ser direcionado a gerir os dados para a reutilização;

- A RDM deve pôr em relevância o conhecimento do valor dos dados, junto com as práticas de proteção e preservação para garantir que este conhecimento e o seu valor, sejam possíveis de serem utilizados na atualidade e em longo prazo;
- Deve-se construir uma estrutura de atenção para uma RDM abrangente no ambiente, para que o valor dos dados seja observado com a devida importância, através de práticas que os identifiquem e os façam conhecidos publicamente dentro e fora da universidade;
- É requerido do investigador que atente com instrumentos necessários para o nível de sensibilidade dos dados, sendo preciso analisar e determinar de modo nítido o que é público e o que é privado, o que é pessoal, confidencial ou sensível;
- Os dados sensíveis devem ser vistos sob a possibilidade de serem ou não partilhados, um cuidado que deve permear a RDM, bem como deve-se atentar aos conflitos decorrentes que interferem na partilha;
- Os instrumentos legais, éticos e as ferramentas tecnológicas e políticas bem definidas, devem ser utilizadas em relação aos dados de natureza sensível, juntamente com as práticas para constituir a proteção legal e técnica, objetivando evitar uma partilha inadequada e permitir a segurança para partilhar os dados necessários; e
- A proteção dos dados deve ser realizada de modo claro, posta em prática através de instrumentos de apoio ao investigador e de acordo com as especificidades dos dados.

É principalmente na perspectiva das práticas essenciais para a RDM que o modelo final acolhe e responde a estas conclusões recomendatórias sobre os dados de investigação. Esta perspectiva pode ser utilizada na caracterização dos dados e na leitura do ambiente onde os dados são gerados e geridos, a partir das etapas de planeamento e descrição, de modo a sugerir onde estes objetos informacionais podem ou poderiam ser reutilizados, também determinando o seu nível de sensibilidade, as informações representativas sobre eles e sobre a proteção e preservação que necessitam, com parâmetros para que os dados e o seu valor, na atualidade e em longo prazo, possam ser preservados, conhecidos, contextualizados e reutilizados. Nesta direção, o modelo possibilita que a universidade construa uma base de atenção e apoio, de modo que o investigador seja direcionado a gerir os dados para a partilha e para a reutilização, mas também para conhecer os dados e o seu valor, permitindo saber o que fazer com eles em termos estruturais, bem como a percepção

de prejuízo das eventuais perdas, ampliando assim a importância e o uso em aspectos valorativos para o emprego de modo público, interdisciplinar e social.

Sob a contribuição da perspectiva dos atores articulados, o modelo indica a construção de uma estrutura de organização e apoio aos dados e a sua gestão, onde as políticas e os instrumentos éticos, legais e as ferramentas tecnológicas são relacionadas entre si, no que diz respeito aos dados de natureza sensível, para que através de práticas para a proteção legal e técnica seja evitada uma partilha inadequada e permita a segurança em partilhar. A necessidade de realizar tais atividades em um nível abrangente, em que os dados heterogêneos sejam geridos na investigação, mas também ampliados para a universidade, têm nas etapas de estruturação e monitoração das práticas essenciais o vínculo requerido, sob uma abordagem multidimensional, entre a investigação local e o institucional global. E ainda, através da perspectiva da metodologia para a adaptação contextual, o modelo permite que os dados sejam geridos para alcançar os itens comuns e mais simples primeiro, seguindo para uma maior complexidade e abrangência institucional, de modo que haja suporte para as UI&D, ao mesmo tempo que permite o conhecimento sobre os dados pela universidade.

As conclusões confrontadas com os parâmetros teóricos, que resultaram nas soluções viabilizadas pelo modelo em relação aos dados a serem geridos, ratificam as ações para uma gestão e partilha de acordo com as especificidades dos dados e do contexto ambiental. Isto é possível porque o modelo considera os dados como o objeto informacional central, o investigador como detentor de informações fundamentais, permitindo que através destes atores sejam viabilizadas uma gestão adaptada ao ambiente, com os atores articulados e práticas necessárias para proporcionar uma partilha adequada dos dados de investigação.

Os atores relacionais nas práticas para a RDM e para a partilha dos dados

Os dados coletados para o estudo relacional dos atores envolvidos nas práticas para a RDM e na partilha dos dados, tiveram perguntas em relação aos seguintes agentes: as pessoas responsáveis; os processos e os fluxos de trabalho; os colaboradores e influenciadores; as ferramentas e serviços; as políticas, as normas e orientações adotadas. Os dados coletados possibilitaram chegar as conclusões recomendatórias para cada ator separadamente, conforme são dispostas a seguir.

Ator - pessoas envolvidas

- É preciso identificar as áreas de responsabilidades e definir as atribuições para as pessoas participantes da RDM, de modo formal e, se possível, compondo a estrutura organizacional, com o objetivo de normatizar o apoio ao investigador para a RDM;
- Deve haver um planeamento e divisão do trabalho em relação as pessoas envolvidas na RDM, considerando as práticas atualizadas das tecnologias informáticas e o contexto ambiental, de modo que permita soluções adequadas para a realidade da investigação e das UI&D, evitando a justaposição de responsabilidades e a sobrecarga de trabalho; e
- O investigador deve participar do planeamento das atribuições e das responsabilidades das pessoas envolvidas na RDM, pois é o investigador quem lida diretamente com os dados e com os vários aspectos para a sua gestão e partilha.

O modelo proposto cumpre as recomendações sobre as pessoas envolvidas na RDM, sobretudo na perspectiva da metodologia para adaptação contextual, que permite identificar, organizar e articular as atribuições, as responsabilidades e a divisão do trabalho para a RDM, as adequando à realidade da investigação. Assim, pode-se atentar para uma possível justaposição e sobrecarga de trabalho, em especial a do investigador. A perspectiva das práticas essenciais, sob a etapa de planeamento, também contribui para o cumprimento das recomendações através da previsão das pessoas necessárias para a RDM. Embora, quando da aplicação do modelo proposto, a necessidade da estruturação e da formalização das atribuições possam ser verificadas e recomendadas, a efetivação do apoio ao investigador na estrutura organizacional da instituição foge do escopo do modelo.

Ator – processos e fluxos de trabalho

Este ator, concernente a organização dos fluxos e processos de trabalho, reúne em si vários outros atores, sendo ele mesmo um elemento importante para a organização da RDM. O objetivo específico produziu para esta parte do trabalho as seguintes conclusões:

- O processo ou fluxo deve ser flexível e adaptado às necessidades e ao perfil da investigação, de modo que contenha práticas comuns, mas que também permita a inserção de novos elementos para a organização do trabalho;

- A necessidade da preservação dos dados pode ter seu direcionamento nos fluxos e processos de trabalho, minimizando as perdas por obsolescência ou desconhecimento;
- Deve-se tratar os fluxos de forma atenta ao processo da investigação e ao comportamento dos participantes, para que a RDM seja realizada com o mínimo de perdas informacionais; e
- Um processo ou fluxo de trabalho sob a filosofia aberta, pode ser pensado no sentido da construção comunitária e colaborativa.

Em relação a flexibilidade, o modelo mesmo sendo constituído por fluxos e processos comuns e padronizados, permite que novos elementos processuais possam ser inseridos para a organização dos trabalhos e adequados ao ambiente, através da perspectiva da metodologia para a adaptação contextual. No mesmo sentido, a perspectiva permite que novos fluxos e processos sejam desenvolvidos de forma que tanto os atores, quanto as comunidades envolvidas, sejam percebidas no processo, possibilitando uma construção comunitária e colaborativa que contribuem com a flexibilidade. A atenção às novidades, também pode ser observada na perspectiva das forças influenciadoras, como as oriundas da filosofia aberta e através do comportamento dos participantes. No fluxo padrão do modelo, a recomendação sobre a minimização das perdas informacionais e a obsolescência das ferramentas, são orientadas pela perspectiva das práticas essenciais, materializada nas etapas de preservação, monitoramento, estruturação e na fase concernente a perda de dados.

Ator – colaboradores e influenciadores

As conclusões recomendatórias seguintes, provêm de atores que podem colaborar e influenciar para a RDM e na partilha, sob vários quadrantes da ciência e da sociedade.

- Esforços devem ser empreendidos para que as práticas e a articulação dos atores colaboradores com a RDM, sejam direcionados de forma que a influência se torne positiva e organizada para a partilha dos dados;
- É necessário o estabelecimento de uma rede colaborativa, de articulações abrangentes que permitam explorar o potencial comum, com a participação dos gestores e da liderança acadêmica da universidade;

- Apresenta-se a necessidade de atores humanos para auxiliar os investigadores em questões que envolvam a RDM no ambiente;
- A literacia, a promoção e o treinamento na educação aberta é importante para suprir as lacunas no conhecimento e proporcionar habilidades em relação a RDM;
- A transversalidade da ciência aberta, dos recursos econômicos e da tradição disciplinar, são fatores de atenção que podem influenciar em aspectos positivos e negativos o investigador para que as colaborações sejam direcionadas em benefício da RDM;
- É necessário estratégias para que os recursos econômicos ou mesmo a sua escassez, sejam administrados para o constante desenvolvimento da RDM na universidade; e
- É preciso atentar para as atualizações de formas e práticas colaborativas no ambiente “aberto”, permitindo adaptar-se a novos processos e tecnologias sob a filosofia aberta.

A influência na colaboração está destacada no modelo, de modo particular, na perspectiva das forças transversais influenciadoras. São atores primordialmente não articulados para a RDM que podem influenciar nesta atividade e nas decisões dos investigadores sobre a gestão e a partilha dos dados. Contemplados no modelo estão os recursos econômicos, as agências de fomento e as novidades colaborativas da ciência e da sociedade sob uma filosofia “aberta”. A percepção das influências está direcionada pela perspectiva da metodologia para a adaptação contextual, identificada e observada para minorar os aspectos negativos e destacar os positivos, de modo que a RDM e a partilha dos dados sejam direcionadas para obter os benefícios destas colaborações influenciadoras.

As recomendações sobre a colaboração dos gestores e a liderança acadêmica da universidade podem ser vistas no mínimo em dois aspectos. O primeiro é na previsão da participação deles sob a construção multivocal para que a RDM seja adaptada ao ambiente. O segundo tipo de colaboração são as decisórias, que independem diretamente do modelo como, por exemplo, o direcionamento para os recursos econômicos, treinamentos e investimentos em educação aberta ou o estabelecimento de uma rede colaborativa que permita explorar o potencial comum da RDM. Os atores humanos para auxiliar os investigadores, em termos de recursos, também estão em dependência decisória, mas podem ser previstos e incluídos na etapa de planeamento, sob a perspectiva das práticas essenciais,

em nível local. Em nível global e na perspectiva dos atores articulados, o auxílio aos investigadores pode ser concebido sob as atribuições de trabalho.

Ator – ferramentas e serviços

As ferramentas e serviços para a RDM e para a partilha dos dados, que remetem de forma geral às infraestruturas, resultaram nas seguintes conclusões e recomendações:

- A infraestrutura organizacional e tecnológica deve ser pensada de modo a melhor aproveitar as ferramentas e serviços, com o uso de propriedades intrínsecas das redes colaborativas e da multidisciplinaridade;
- As infraestruturas devem ser percebidas sob as características de um conhecimento comum, com recursos partilhados, observadas sob a perspectiva de que as pessoas estão sujeitas a dilemas sociais e sob tensões abundantes entre as partes interessadas;
- A governança da universidade necessita direcionar investimentos financeiros, políticos e de capacitação para tornar as infraestruturas articuladas, interoperáveis e sustentáveis, mas também para minorar tensões e conflitos decorrentes;
- É necessário o acompanhamento dos recursos para que a articulação com outros atores possa ser realizada de forma comum e de modo a permitir a interoperabilidade em múltiplos sistemas e ferramentas de informação, no âmbito da instituição e fora dela;
- As práticas sustentáveis devem estar diretamente ligadas às ferramentas, serviços e ações de governança, para que proporcionem a sustentabilidade;
- É pertinente envolver equipes técnicas de várias áreas juntamente com a governança da universidade, em um arranjo lógico e sob igual importância, para abordar questões amplas como as estruturas a nível da universidade ou mais específicas como a usabilidade dos materiais;
- É preciso dimensionar e articular as infraestruturas tecnológicas e organizacionais, além de determinar o grau de investimento na preservação em longo prazo, com uso e construção de políticas, estratégias e ferramentas direcionadas para os cuidados com os dados;
- O planeamento para delinear a infraestrutura de como os dados serão armazenados e como o acesso e a segurança serão gerenciados, deve ser uma prática para a organização e para a minimização do risco de perdas informacionais; e

- É necessário que as ferramentas e serviços sejam trabalhadas com outras competências, de modo que a RDM seja realizada sob uma infraestrutura não somente ferramental e tecnológica, mas esclarecedora da importância relacional para os seus usuários.

O modelo proposto tem nas ferramentas e serviços um ator a ser estruturado e ao mesmo tempo estruturante. No primeiro caso, a dimensão do ferramental pode ser concebida no modelo através da perspectiva das práticas essenciais e da metodologia para a adaptação contextual, onde a própria metodologia e as etapas de planeamento, monitoramento e principalmente de estruturação, são direcionadas de modo a garantir tanto a concepção, quanto o acompanhamento para manter as ferramentas e serviços a funcionar de modo adequado, incluindo as questões sobre a interoperabilidade em múltiplos sistemas e ferramentas no âmbito da instituição e fora dela, bem como a sustentabilidade diretamente ligadas a tais componentes. O envolvimento de equipes técnicas de várias áreas e da governança da universidade é previsto na metodologia, podendo resultar no correto dimensionamento e na articulação das infraestruturas tecnológicas e organizacionais, para que possam funcionar em redes colaborativas e multidisciplinares objetivadas para o conhecimento comum e para a partilha de recursos. Estas perspectivas proporcionam também planejar e indicar o grau de investimento na preservação, na construção de políticas, estratégias e meios direcionadas para os cuidados com os dados da investigação.

No segundo caso, o modelo pode atuar de modo estruturante para a RDM através da implementação das recomendações apontadas nestas conclusões, especialmente no que se refere as ferramentas e serviços, pois possibilitam diminuir as diferenças ambientais e barreiras com o uso de recursos partilhados. Para esta possível implementação é necessário observar as tensões e conflitos próprios do ambiente, bem como as questões essenciais para a interoperabilidade e para a sustentabilidade dos elementos que permeiam a RDM, a exemplo da preservação e da utilização dos dados em longo prazo. A atuação estruturante também é viabilizada por intermédio das práticas essenciais, através de etapas como a “preservação”, que possui a capacidade de prover o apoio nas questões físicas e lógicas para o armazenamento, curadoria, partilha e reutilização dos dados. Mas também através da atenção direta no ator “ferramentas”, que em ligação com as práticas essenciais e a perspectiva das forças influenciadoras, contém o que é necessário para a RDM estabelecer

o relacionamento da infraestrutura tecnológica e dos serviços com os outros atores do modelo. Isto é possível sob a direção da metodologia para a adaptação contextual, que conduz soluções participativas, sob vários níveis, para que a RDM seja adaptada à realidade do ambiente, observando as ferramentas e serviços no contexto local e global.

No entanto, as conclusões que apontam para a implementação de uma estrutura de ferramentas e serviços partilhados, bem como o direcionamento dos investimentos financeiros, políticos e de capacitação, devem ser planeados e direcionados. Neste sentido, ratifica-se que, para que estas recomendações possam ser cumpridas, o envolvimento dos atores em níveis de decisão da universidade é primordial. Portanto, são recomendações que podem ser previstas, mas ultrapassam o escopo de realização do modelo para a RDM.

Ator – políticas, normas e orientações

Os dados coletados que resultaram nas conclusões abaixo, referem-se às documentações que direcionam os dados para serem geridos, cuidados e partilhados.

- As políticas, regras e documentos existentes ou a serem elaborados que envolvem a RDM, devem ser prioridade para orientar as práticas que possibilitam a gestão e a partilha dos dados na universidade;
- As políticas devem ser organizadas e implementadas para funcionar como um agregador de princípios e práticas que auxiliam os investigadores a lidar com os dados;
- As questões legais devem ser organizadas nos níveis global e local, de modo que sejam complementares e não sobrepostas;
- A construção de uma proteção legal para a RDM, deve ser direcionada para diminuir as vulnerabilidades e proporcionar segurança no contexto da investigação e dos atores;
- Os metadados são obrigatórios como uma prática para assegurar que a RDM possa ser realizada com a finalidade de que os dados sejam reutilizáveis;
- É necessário dispor das condições técnicas e infraestruturais, de modo que o investigador possa ser apoiado por meios articulados e que os metadados sejam criados a contento das especificidades da investigação;

- As UI&D, as investigações e os investigadores devem ser preparados através de políticas, infraestrutura, esclarecimentos e práticas para que a elaboração de um plano de gestão de dados seja recebido de modo a gerar o mínimo de barreiras; e
- A biblioteca é um ator importante que pode colaborar com as práticas para a RDM em termos de documentação de padrões e metadados.

As políticas, normas e orientações documentais existentes e a serem elaboradas para a RDM estão priorizadas no modelo através das perspectivas da articulação dos atores, das práticas essenciais e da adaptação contextual. Esta última permite a análise para identificar e avaliar o que existe, mas também a necessidade da criação de políticas, verificando as normativas das investigações, as políticas nacionais, internacionais e da própria universidade. Na perspectiva da articulação e através do ator “políticas”, permite-se que sejam organizados os normativos necessários que possam resultar em um indicador documental convergente, que clarifica o uso das políticas e evita sobreposições, viabilizando as condições técnicas e infraestruturais de modo que o investigador possa ser apoiado para realizar as práticas essenciais para a RDM em uma partilha adequada dos dados. Assim, este ator também pode funcionar como um agregador de princípios e práticas que auxilia os investigadores a lidarem com diversos tipos de questões documentais para as políticas e normativas, bem como a necessária proteção legal decorrente, proporcionando segurança e a minimização das vulnerabilidades e tensões para a partilha dos dados, principalmente sob o acesso aberto.

Na perspectiva das práticas essenciais, o modelo permite tratar das políticas, normas ou orientações documentais em todas as suas etapas. Em destaque, na fase da descrição, contempla-se os metadados como uma prática para assegurar que a RDM torne os dados reutilizáveis. Está também prevista na etapa de planeamento, a elaboração de um plano de gestão de dados que repercute em todo o processo para a RDM. As outras perspectivas do modelo também interagem sob a orientação das políticas, seja na articulação com as comunidades envolvidas, nas atribuições de trabalho ou nas ferramentas utilizadas, assim como na possibilidade da interferência colaborativa na perspectiva das forças influenciadoras. Referenciado nestas conclusões recomendatórias como um ator importante, a previsão da participação da biblioteca na colaboração com os padrões e metadados está inserida sob a perspectiva das forças transversais influenciadoras.

As práticas para a partilha dos dados em acesso aberto

O resultado da percepção dos investigadores das UI&D, em relação aos hábitos para a partilha dos dados em acesso aberto, produziram as seguintes conclusões:

- O ato da partilha requer uma série de práticas anteriores, deve ser visto em conjunto e em contexto com o ambiente da investigação e do investigador, de forma a garantir os benefícios que culminam da reutilização dos dados;
- Há a necessidade em saber onde estão os dados já partilhados;
- Deve-se atentar para os motivos da não partilha dos dados, de modo a percebê-los com um trabalho volumoso, que requer o envolvimento de vários atores para a minimizá-los;
- É necessário considerar, de modo organizado e articulado, as motivações como elementos propulsores para a partilha dos dados em acesso aberto;
- Deve-se trabalhar para minimizar as barreiras, de modo que as motivações negativas sejam identificadas e as positivas sejam postas em relevo;
- As motivações negativas e positivas não podem ser vistas isoladamente, pois estão intimamente relacionadas entre si e ao contexto; e
- As soluções devem ser pensadas de modo global a partir das práticas locais.

As conclusões sobre a partilha dos dados em acesso aberto estão abrangidas no modelo para a RDM, pois tal modelo é todo direcionado para uma partilha adequada dos dados, que em sua definição deve ser sempre que possível em acesso aberto e de modo a permitir a reutilização dos dados. Consequentemente, as perspectivas do modelo são orientadas para que este tipo de partilha seja realizada. No entanto, é na perspectiva da metodologia para a adaptação contextual que o uso associado das perspectivas do modelo são direcionadas para a realização da partilha dos dados em acesso aberto, de forma global, a partir de práticas locais, possibilitando uma visão em conjunto e em contexto com a universidade, com o ambiente da investigação e do investigador. É nesse sentido que o ato da partilha deve ser visto em conjunto, envolvendo os atores articulados e as práticas essenciais requeridas para a RDM, onde a interrelação entre estes e as demais perspectivas, possibilitam os elementos necessários para que uma partilha adequada seja realizada. Assim, mesmo os dados já partilhados em acesso aberto podem ser conhecidos.

As motivações para que a partilha seja realizada em acesso aberto, podem ser verificadas através das interferências das forças transversais influenciadoras. Essas motivações e tais forças, percebidas com o uso da metodologia para a adaptação contextual, são propulsoras para uma partilha organizada, articulada e adequada, pois o modelo permite que as motivações negativas e positivas sejam identificadas, postas em relevo e organizadas com a finalidade de serem pensadas de modo global, através das articulações dos atores, para repercutir de modo local, nas práticas essenciais para a gestão. Assim, o modelo deve possibilitar que as barreiras sejam minimizadas e os benefícios potencializados para que a RDM proporcione uma partilha adequada dos dados.

A proposição de um modelo para a RDM, que possibilite uma partilha adequada dos dados, tendo por base a articulação dos atores e o uso de práticas essenciais para a gestão, no ambiente da universidade: foi o objetivo geral aqui cumprido, pois o modelo apresentado proporciona que no ambiente da universidade, observado na perspectiva das UI&D, os dados produzidos possam ser geridos a partir das investigações, através da articulação dos atores e por intermédio de práticas essenciais, ambas objetivadas para a realização de uma gestão e partilha adequada dos dados.

Mais especificamente, o modelo para a RDM alcançou o objetivo proposto porque possibilita que tanto a articulação dos atores, quanto as práticas essenciais estejam interligadas para que a partilha dos dados seja assegurada por meio de infraestruturas digitais e se possível, na forma do acesso aberto, com todo o processo direcionado para a reutilização dos dados. Esta interligação permite que os níveis global e local estejam representados e conexos, de modo que o institucional conheça os dados e colabore com a investigação e a investigação influa e seja apoiada pela instituição, viabilizado no modelo principalmente na perspectiva da metodologia para a adaptação contextual.

Aplicada de forma organizada e através de uma construção multivocal, a metodologia para a adaptação contextual analisa e efetiva todas as perspectivas do modelo, bem como caracteriza os dados e as especificidades do ambiente, considerando também nesse percurso as forças influenciadoras que podem interferir na gestão e na partilha dos dados de investigação. O resultado é uma RDM adaptada ao ambiente com a ligação entre o local e o institucional, permitindo que os dados geridos, sejam transformados em dados abertos e reutilizáveis.

No sentido geral, o modelo para a gestão dos dados de investigação: propõe articulações e práticas para a partilha dos dados na universidade. Objetivamente, a potencialização deste modelo para a RDM, através da articulação dos seus atores e no uso de práticas essenciais para a gestão dos dados no ambiente das UI&D na universidade, resulta em uma partilha adequada dos dados. É o produto alcançado por esta investigação.

Contribuições do estudo

As universidades estão sob o desafio de adaptarem-se a uma nova realidade para lidar com a produção massiva, com o valor e com a importância dos dados resultantes das atividades da investigação. A RDM é responsável pelo tratamento dos dados que alimenta e procede da investigação realizada no âmbito da universidade. Por esta razão, uma RDM eficiente, resultado da articulação dos seus atores e da realização de práticas essenciais para a gestão e partilha adequada dos dados, influencia diretamente na atividade e no desempenho da investigação, dos investigadores, das UI&D e da própria universidade.

Quanto mais clara, organizada e apoiada for a RDM, melhor os investigadores realizarão suas atividades, visto que tais atividades também têm mudado significativamente, direcionadas pela necessidade dos dados serem partilhados em acesso aberto. Somados a isto, as UI&D e a UC podem organizar a RDM para que os investimentos sejam realizados de modo gradual, planeado e coordenado, considerando as transformações e a complexidade da atividade, de modo a conhecer os dados oriundos da investigação e a explorar as potencialidades destes dados a favor da própria instituição, da ciência e da sociedade.

Entretanto, como apresentado ao longo deste trabalho, são inúmeros os desafios do ambiente da RDM na universidade em que se inserem as UI&D. Tendo em vista estes desafios, o modelo para a RDM está proposto com a finalidade de atender de modo efetivo às demandas emergentes do ambiente, observando com a devida importância os problemas de organização dos atores envolvidos, da fragmentação e da invisibilidade dos dados, materializado pelas práticas isoladas e em consequência da complexidade e da heterogeneidade que caracterizam os dados e o ambiente da investigação na universidade.

A solução, como proposto, passa pela conjunção organizada das perspectivas da articulação dos atores, das práticas essenciais para gerir e partilhar os dados de investigação de

forma adequada, considerando neste percurso as influências sociotécnicas e a necessária adaptação da RDM ao ambiente. Assim, é permitido dar visibilidade aos dados, através de atores institucionais coordenados, ligados a práticas comuns na investigação, que resultam na partilha dos dados objetivadas ao acesso aberto, de modo a possibilitar um conhecimento organizado sobre os dados através da RDM, além de tornar possível a reutilização dos dados que já foram partilhados. É importante ressaltar que nenhuma das perspectivas do modelo dispõe, de forma isolada, de instrumental teórico e metodológico para tratar os problemas discutidos.

Portanto considera-se que o modelo proposto, além de um avanço científico no reconhecimento de como tais desafios podem ser observados e solucionados é também original, inovador e constitui uma estrutura bem delineada, com um conjunto sólido de fundamentos norteadores para que uma RDM seja implementada sob a articulação dos atores e realizada com práticas essenciais de gestão no âmbito da universidade, resultando na partilha adequada dos dados e nos benefícios decorrentes relatados nesta investigação, sob a área da ciência aberta.

Limitações da investigação

As limitações da investigação tiveram seu lugar primeiramente na impossibilidade da inclusão da área das ciências exatas dentre os dados coletados, de acordo com o domínio científico adotado pelo Instituto de Investigação Interdisciplinar. Além disso, a pandemia causada pela COVID-19, inviabilizou a investigação de forma presencial, que foi assim iniciada e interrompida pelo fechamento temporário da Universidade de Coimbra.

Sugestões para estudos futuros

Sugerir uma agenda que estimule futuras investigações e desdobramentos do que aqui foi produzido é um exercício desafiador, posto que a temática é recente e aborda questões relativamente mutáveis. No entanto, é pertinente realizar tais sugestões pois há muitas questões que podem ser desenvolvidas e atualizadas a partir das abordagens aqui realizadas. Na realidade, um dos intentos deste trabalho e do modelo resultante, é proporcionar o início de discussões e investigações nas diversas vertentes abertas e não exploradas nesta tese, bem como a alteração e expansão do próprio modelo, a ser desenvolvido também através da área multidisciplinar da Ciência da Informação. Assim, como sugestão para estudos futuros, pode-se pensar nas seguintes abordagens:

- O detalhamento de cada etapa das práticas essenciais, para gerar um ferramental estruturado que sirva de apoio aos utilizadores da RDM, a partir do modelo proposto;
- A indicação de outras forças influenciadoras ou seu agrupamento, de modo a especificar como tais elementos interferem no contexto a ser aplicado;
- A maneira específica como a universidade viabilizaria uma RDM comum às UI&D, com os atores articulados e de modo a partilhar os recursos humanos e tecnológicos;
- O detalhamento da realização da metodologia para a adaptação contextual, no sentido de especificar minuciosamente cada etapa do método de intervenção;
- A ênfase na literacia para a aplicação do modelo e no que envolve a RDM;
- O estudo do modelo a partir de temas relevantes da área, a exemplo dos princípios FAIR e das metodologias colaborativas sob a filosofia aberta como as do tipo “GitHub”;
- O acréscimo de conhecimento, sob a abordagem da *knowledge infrastructure* e o impacto deste conhecimento acrescido sobre o modelo;
- A implementação gradual do modelo, a partir das UI&D até o alcance de todos os ambientes da universidade, através de uma rede de sistemas de atividades heterogêneas e complexas, mas também entre universidades ou em um ambiente global; e
- A validação do modelo junto aos investigadores, gestores da universidade e profissionais da informação, incluindo as partes interessadas que atuam para a RDM e para a partilha dos dados no ambiente.

REFERÊNCIAS

- Abbott, D. (2008). *DCC Briefing Paper: What is digital curation?* Era.ed.ac.uk. <https://era.ed.ac.uk/handle/1842/3362>
- Achard, P., Ayris, P., Fdida, S., Gradmann, S., Horstmann, W., Labastida, I., Lyon, L., Maes, K., Reilly, S., Smit, A. (2013). *LERU Roadmap for research data (LERU Advice Paper, 14)*. Recuperado em 12 de novembro de 2021, de <https://www.leru.org/files/LERU-Roadmap-for-Research-Data-Full-paper.pdf>
- Albagli, S., Maciel, M. L., & Abdo, A. H. (Orgs.). (2015). *Ciência aberta, questões abertas*. IBICT. <http://livroaberto.ibict.br/handle/1/1060>
- Albagli, S., Clinio, A., & Raychtock, S. (2014). Ciência aberta: correntes interpretativas e tipos de ação. *Liinc em Revista*, 10(2), 434-450. <https://doi.org/10.18617/liinc.v10i2.749>
- Allen, B. L. (1991). Cognitive research in information science: implications for design. *Annual review of information science and technology*, 26, 3-37. Recuperado em 19 de outubro de 2021, de <https://www.learntechlib.org/p/145403/>
- Antelman, K. (2004). Do Open-Access Articles Have a Greater Research Impact? *College & Research Libraries*, 65(5), 372–382. <http://eprints.rclis.org/archive/00002309/>
- Antelman, K. (2006). Self-archiving practice and the influence of publisher policies in the social sciences. *Learned Publishing*, 19(2), 85-95. <https://doi.org/10.1087/095315106776387011>
- Appelbe, B., & Bannon, D. (2007). eResearch – Paradigm shift or propaganda? *Journal of Research and Practice in IT*, 39(2), 83–90. <https://search.informit.org/doi/10.3316/informit.937758621242625>
- Appolinário, F. (2007). *Dicionário de metodologia científica: um guia para a produção do conhecimento científico*. Atlas. Recuperado em 13 de agosto de 2021, de <https://pesquisa.bvsalud.org/portal/resource/pt/lil-591596>
- Arabito S., & Pitrelli N. (2015). Open science training and education: Challenges and difficulties on the researchers' side and in public engagement. *Journal of Science Communication*, 14 (4). <https://doi.org/10.22323/2.14040301>
- Asher, A.D., Deards, K.D., Esteva, M., Halbert, M., Jahnke, L.M., Jordan, C., Keralis, S., Kulasekaran, S., Moen, W.E., Stark, S., Urban, T., & Walling, D. (2013). *Research Data Management Principles, Practices, and Prospects*. Recuperado em 15 de setembro de 2021, de <https://digital.library.unt.edu/ark:/67531/metadc234929/>
- Autran, M. M., & Borges, M. M. (2014). Comunicação da ciência: (r)evolução ou crise?, *RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde*, 8(2), 122-138. <http://hdl.handle.net/10316/26596>

- Ayris, P., Bernal, I., Cavalli, V., Dorch, B., Frey, J., Hallik, M., ... & Horstmann, W. (2018). *LIBER Open Science Roadmap*. <https://doi.org/10.5281/zenodo.1303002>
- Ball, A. (2010). *Review of the State of the Art of the Digital Curation of Research Data*. University of Bath. Recuperado em 09 de dezembro de 2021, de <https://researchportal.bath.ac.uk/files/293012/erim1rep091103ab12.pdf>
- Ball, A. (2012). *Review of Data Management Lifecycle Models*. University of Bath. Recuperado em 09 de dezembro de 2021, de <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>
- Banović, J. (2021). The Future of Science. *Infotheca - Journal for Digital Humanities*, 20(1-2), 47-55. https://doi.org/10.18485/infotheca.2020.20.1_2.3
- Bardin, L. (2011). *Análise de conteúdo*. Edições 70.
- Bartling, S., & Friesike, S. (2014). Towards another scientific revolution. In S. Bartling, & S. Friesike (Eds.), *Opening science*, 3–15. Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_1
- Beagrie, N. (2004). The Continuing Access and Digital Preservation Strategy for the UK Joint Information Systems Committee (JISC). *D Lib Mag.*, 10(7/8). Recuperado em 09 de dezembro de 2021, de <http://www.dlib.org/dlib/july04/beagrie/07beagrie.html>
- Beagrie, N. (2006). Digital Curation for Science, Digital Libraries, and Individuals. *Int. J. Digit. Curation*, 1, 3-16. <https://doi.org/10.2218/ijdc.v1i1.2>
- Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179-179. <https://doi.org/10.1038/489179a>
- Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., ... & Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2), 599-611. <https://doi.org/10.1016/j.future.2011.08.004>
- Bernal, J. D. (1939). Social function of science. Recuperado em 03 de dezembro de 2021, de <https://bit.ly/2W2OpkG>.
- Bertalanffy, L. V. (1973). *Teoria geral dos sistemas*. Editora Vozes.
- Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes P., Görögh, E., ... & Heller, L., (2018). *Open Science Training Handbook*. <https://doi.org/10.5281/zenodo.1212496>
- Bezuidenhout, L. M., Leonelli, S., Kelly, A. H., & Rappert, B. (2017). Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy*, 44(4), 464-475. <https://doi.org/10.1093/scipol/scw036>
- Björk, B. C., & Solomon, D. (2014). *Developing an Effective Market for Open Access Article Processing Charges*. <https://doi.org/10.5281/zenodo.51788>
- Black, M. (1962). *Models and metaphors: Studies in language and philosophy*. Cornell University Press.

- Borges, M. M. (2006). *A esfera: Comunicação académica e novos media* [Tese de Doutoramento, Universidade de Coimbra]. <http://hdl.handle.net/10316/8557>
- Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press. <https://doi.org/10.7551/mitpress/7434.001.0001>
- Borgman, C. L. (2009). The Digital Future is Now: A Call to Action for the Humanities. *Digital Humanities Quarterly*, 3(4). Recuperado em 07 de dezembro de 2021, de <https://escholarship.org/uc/item/0fp9n05s>
- Borgman, C. L. (2010). Research Data: Who Will Share What, with Whom, When, and Why? *RatSWD Working Paper*. <http://dx.doi.org/10.2139/ssrn.1714427>
- Borgman, C. L. (2010b). The Digital Archive: The Data Deluge arrives in the Humanities. *UCLA: Center for Knowledge Infrastructures*. Recuperado em 02 de dezembro de 2021, de <https://escholarship.org/uc/item/2cc978tf>
- Borgman, C. L. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 63(6): 1059-1078. Recuperado em 02 de dezembro de 2021, de <https://ssrn.com/abstract=2073876>
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.
- Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, 33(2). Recuperado em 03 de dezembro de 2021, de <https://escholarship.org/uc/item/9dh6b8j0>
- Borgman, C. L., & Bourne, P. E. (2022). *Why it takes a village to manage and share data*. <https://doi.org/10.48550/arXiv.2109.01694>
- Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: data, diversity, and digital libraries. *International Journal on Digital Libraries*, 16(3), 207-227. <https://doi.org/10.1007/s00799-015-0157-z>
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2018). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *UCLA: Center for Knowledge Infrastructures*. Recuperado em 12 de dezembro de 2021, de <https://escholarship.org/uc/item/9qv3x7cx>
- Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, 7(1), 17-30. <https://dl.acm.org/doi/10.5555/1296748.1296756>
- Borgman, C. L., Golshan, M.S., Sands, A.E., Wallis, J.C., Cummings, R.L., Darch, P.T., & Randles, B.M. (2016). Data management in the long tail: Science, software, and service. *International Journal of Digital Curation*, 11(1), 128– 149. <https://doi.org/10.2218/ijdc.v11i1.428>.

- Brackett, M., & Earley, P. S. (2009). *The DAMA guide to the data management body of knowledge (DAMA-DMBOK guide), first edition*. Technics Publications. Recuperado em 12 de novembro de 2021, de https://www.academia.edu/19992490/The_DAMA_Guide_to_the_Data_Management_Body_of_Knowledge_First_Edition
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, 32(7), 513. <https://psycnet.apa.org/doi/10.1037/0003-066X.32.7.513>
- Brown, S., Bruce, R., & Kernohan, D. (2015). *Directions for Research Data Management in UK Universities*. <http://repository.jisc.ac.uk/5951/>
- Bryant, R., Lavoie, B., & Malpas, C. (2017). A Tour of the Research Data Management (RDM) Service Space. *The Realities of Research Data Management, Part 1*. OCLC Research. <https://doi.org/10.25333/C3PG8J>.
- Budapest Open Access Initiative. (2002). *Read Budapest Open Access Initiative*. Recuperado em 12 de dezembro de 2021, de <http://www.budapestopenaccessinitiative.org/read>.
- Budroni, P., Claude-Burgelman, J., & Schouppe, M. (2019). Architectures of knowledge: the European open science cloud. *ABI Technik*, 39(2), 130-141. <https://doi.org/10.1515/abitech-2019-2006>
- Burt, P. V., & Kinnucan, M. T. (1990). Information models and modeling techniques for information systems. *Annual review of information science and technology*, 25, 175-208. Recuperado em 10 de janeiro de 2022, de <https://api.semanticscholar.org/CorpusID:63443527>
- Callon, M. (1986). The Sociology of an Actor-Network: The Case of the Electric Vehicle. In: Callon, M., Law, J., Rip, A. (eds) *Mapping the Dynamics of Science and Technology*. https://doi.org/10.1007/978-1-349-07408-2_2
- Capra, F. (1982). *O Tao da Física*. Cultrix.
- Carlson, J. (2014). The Use of Life Cycle Models in Developing and Supporting Data Services. In J. M. Ray (Ed.), *Research Data Management: Practical Strategies for Information Professionals*. Purdue University Press. <https://doi.org/10.2307/j.ctt6wq34t.6>
- Carlson, S. (2006). *Lost in a Sea of Science Data* [Review of *Lost in a Sea of Science Data*]. The Chronicle of Higher Education. Recuperado em 15 de janeiro de 2021, de <https://www.chronicle.com/article/lost-in-a-sea-of-science-data/>
- Consultative Committee for Space Data Systems. (2002). Consultative Committee for Space Data Systems - Reference Model for an Open Archival Information System (OAIS). *Recommendation for Space Data System Standards*. Recuperado em 21 de janeiro de 2021, de <http://public.ccsds.org/publications/RefModel.aspx>.
- Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC bioinformatics*, 12 (Suppl 15), S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>

- Chorley, R., & Haggett, P. (2013). *Integrated Models in Geography (Routledge Revivals)*. (1st ed.). Routledge. <https://doi.org/10.4324/9780203075746>
- Choudhury, G.S., Huang, C., & Palmer, C.L. (2020). Updating the DCC Curation Lifecycle Model. *Int. J. Digit. Curation*, 15, 1-12. <https://doi.org/10.2218/ijdc.v15i1.721>
- Clark, M. P., Luce, C. H., AghaKouchak, A., Berghuijs, W., David, C. H., Duan, Q., ... & Tyler, S. W. (2021). Open science: Open data, open models, ...and open publications? *Water Resources Research*, 57(4). <https://doi.org/10.1029/2020WR029480>
- Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Harvard University Press.
- Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. *Distributed cognitions: Psychological and educational considerations*, 1-46. Cambridge University Press.
- Comissão Europeia. (2016). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 Version 2.11* Recuperado em 12 de dezembro de 2021, de <https://www.fosteropenscience.eu/content/guidelines-open-access-scientific-publications-and-research-data-horizon-2020>.
- Comissão Europeia. (2016a). *Open Innovation, Open Science, Open to the World – a vision for Europe*. Recuperado em 10 de maio de 2021, de <https://op.europa.eu/en/publication-detail/-/publication/3213b335-1cbc-11e6-ba9a-01aa75ed71a1>
- Comissão Europeia. (2016b). *Realising the European Open Science Cloud - First report and recommendations of the Commission high level expert group on the European open science cloud*. Recuperado em 10 maio de 2021, de <https://op.europa.eu/s/u8Qr>
- Comissão Europeia. (2020). *Shaping Europe's Digital Future*. Recuperado em 11 mai. 2021, de https://ec.europa.eu/info/sites/default/files/communication-shaping-europes-digital-future-feb2020_en_4.pdf
- Constantopoulos, P., & Dallas, C. (2010). Aspects of a digital curation agenda for cultural heritage. *IEEE International Conference on Distributed Human-Machine Systems*. Recuperado em 09 de dezembro de 2021, de <http://www.dcu.gr/wp-content/uploads/2016/10/Aspects-of-a-digital-curation-agenda-for-cultural-heritage.pdf>
- Cooke, J. (2007). A visual arts perspective on open access institutional repositories. *Proceedings of 23 Computers and History of Art Annual Conference*. Birkbeck 8-9th November. Recuperado em 11 mai. 2021, de <http://research.gold.ac.uk/140/>
- Cooper, D. R., & Schindler, P. S. (2016). *Métodos de Pesquisa em Administração-12ª edição*. McGraw Hill Brasil.
- Corti, L. (2014). *Managing and Sharing Research Data*. Recuperado em 09 de dezembro de 2021, de <https://pdfs.semanticscholar.org/4ee5/933d8a9f4066af55a88696d18a7e8332da82.pdf>

- Corti, L., Van den Eynden, V., Bishop, L., & Morgan Brett, B. R. (2011). *Managing and Sharing Data; Training Resources*. Recuperado em 09 de dezembro 2021, de <http://repository.essex.ac.uk/2398/1/TrainingResourcesPack.pdf>
- Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: a guide to good practice*. Sage Publishing. <https://doi.org/10.25607/OBP-1540>
- Costa, M., & Leite, F. (2017). *Repositórios institucionais da América Latina e o acesso aberto à informação científica*. <http://hdl.handle.net/20.500.11959/brapci/187961>.
- Cougo, P. (2013). *Modelagem conceitual e projeto de banco de dados*. Elsevier Brasil.
- Cox, A.M., & Verbaan, E. (2018). *Exploring Research Data Management*. Facet. <https://doi.org/10.29085/9781783302802>
- Cox, A.M., & Tam, W. (2018) A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, 70(2), 142-157. <https://doi.org/10.1108/AJIM-11-2017-0251>
- Cox, A.M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299-316. <https://doi.org/10.1177/0961000613492542>
- Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. *The Journal of Academic Librarianship*, 37(6), 460–470. <https://dx.doi.org/10.1016/j.acalib.2011.07.002>
- Da Silva, A. M. (2010). Modelos e Modelizações em Ciência da Informação: O Modelo eLit.pt e a investigação em literacia informacional. *Prisma. com*, (13), 298-353. <http://hdl.handle.net/20.500.11959/brapci/71982>
- Dallmeier-Tiessen, S., Darby, R., Goerner, B., Hyppoelae, J., Igo-Kemenes, P., Kahn, D., Lambert, S.C., Lengenfelder, A., Leonard, C., Mele, S., Nowicka, M., Polydoratou, P., Ross, D., Ruiz-Perez, S., Schimmer, R., Swaisland, M., & Stelt, W.V. (2011). Open access journals-what publishers offer, what researchers want. *Information services & use*, 31(1-2), 85-91. <http://dx.doi.org/10.3233/ISU-2011-0624>
- Daniels, H. (2001). Activity theory and knowledge production: Twin challenges for the development of schooling for pupils who experience EBD. *Emotional and Behavioural Difficulties*. 6(2), 113-124. <https://doi.org/10.1080/13632750100507661>
- Daniels, P. J. (1986). Cognitive models in information retrieval—an evaluative review. *Journal of documentation*, 42(4), 272-304. <https://doi.org/10.1108/eb026797>
- Data Documentation Initiative (2021). *DDI Lifecycle 3.3*. Recuperado em 09 de dezembro de 2021, de <https://ddialliance.org/Specification/DDI-Lifecycle/3.3/>
- Data Observation Network for Earth. (2021). *Data Life Cycle*. Recuperado em 09 de dezembro de 2021, de <https://old.dataone.org/data-life-cycle>
- Data Observation Network for Earth. (2021a). *About DataOne*. Recuperado em 09 de dezembro de 2021, de <https://www.dataone.org/about>

- Data Observation Network for Earth. (2021b). *What DataOne*. Recuperado em 09 de dezembro de 2021, de <https://old.dataone.org/what-dataone>
- David, P. (2008). The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. *Capitalism and Society*, 3(2). <https://doi.org/10.2202/1932-0213.1040>
- Davydov, V. V. (1988). The Concept of Theoretical Generalization and Problems of Educational Psychology. *Studies in Soviet Thought*, 36(3), 169–202. Recuperado em 09 de dezembro de 2021, de <http://www.jstor.org/stable/20100369>
- Davydov, V. V., & Radzikhovskii, L. A. (1985). Vygotsky's theory and the activity-oriented approach in psychology. *Culture, communication, and cognition: Vygotskian perspectives*. Cambridge University Press.
- Demo, P. (1985). *Introdução da metodologia*. Atlas.
- Digital Curation Centre. (2021). *What is digital curation?* Recuperado em 09 de dezembro de 2021, de <https://www.dcc.ac.uk/about/digital-curation>
- Digital Curation Centre. (2021a). *DCC Curation Lifecycle Model*. Recuperado em 09 de dezembro de 2021, de <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>
- Ditadi, C. (2003). *Preservação de documentos eletrônicos*. Arquivo Nacional/CTDE.
- Dunning, A., de Smaele, M., Böhmer, J., (2017). Are the fair data principles fair? *International Journal of digital curation*, 12(2), 177-195. <https://doi.org/10.2218/ijdc.v12i2.567>
- Eco, U. (1977). *Como se faz uma tese*. Perspectiva. Recuperado em 09 de dezembro de 2021, de https://edisciplinas.usp.br/pluginfile.php/5755876/mod_resource/content/1/ECO%20C%20Humberto.%20Como%20se%20Faz%20uma%20Tese.%20Perspectiva%20C%2013ed%20C%201977.pdf
- Eilam, G. (2003). The Philosophical Foundations of Aleksandr R. Luria's Neuropsychology. *Science in Context*, 16(4), 551-577. <https://doi.org/10.1017/S0269889703000966>
- e-Infrastructure Reflection Group. (2016). *Long Tail of Data e-IRG Task Force Report*. Recuperado em 01 de fevereiro de 2021, de <http://eirg.eu/documents/10920/238968/LongTailOfData2016.pdf>
- Elliott, K. C., & Resnik, D. B. (2019). Making open science work for science and society. *Environmental health perspectives*, 127(7). <https://doi.org/10.1289/EHP4808>
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of education and work*, 14(1), 133-156. <https://www.tandfonline.com/doi/abs/10.1080/13639080020028747>
- Engeström, Y. (2004). New forms of learning in co-configuration work. *Journal of Workplace learning*. <https://doi.org/10.1108/13665620410521477>

- Engeström, Y. (2014). *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139814744>
- Engeström, Y. (2016). *Studies in Expansive Learning: Learning What Is Not Yet There*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316225363>
- Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational research review*, 5(1), 1-24. <https://doi.org/10.1016/j.edurev.2009.12.002>
- Engeström, Y., & Sannino, A. (2011). Discursive manifestations of contradictions in organizational change efforts: A methodological framework. *Journal of Organizational Change Management*, 24(3), 368–387. <https://doi.org/10.1108/095348111111132758>
- Engeström, Y., & Sannino, A. (2012). Whatever happened to process theories of learning? *Learning, Culture and Social Interaction*, 1(1), 45-56. <https://doi.org/10.1016/j.lcsi.2012.03.002>
- Faniel, I. M. & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for largescale data sharing and reuse. *International Journal of Digital Curation*, 6(1), 58-69. <https://doi.org/10.2218/ijdc.v6i1.172>
- Fundação para a Ciência e Tecnologia. (2021). *Unidades de I&D*. Recuperado em 14 de Junho de 2021, de <https://www.fct.pt/apoios/unidades/index.phtml.pt>
- Fecher B., & Friesike S. (2014) Open Science: One Term, Five Schools of Thought. In: Bartling S., Friesike S. (eds) *Opening Science*, 17-47. https://doi.org/10.1007/978-3-319-00026-8_2
- Ferguson, A.R., Nielson, J.L., Cragin, M.H., Bandrowski, A.E., & Martone M.E. (2014). Big data from small-data: data sharing in the ‘long tail’ of neuroscience. *Nature Neuroscience*, 17(11), 1442–1447. <https://doi.org/10.1038/nn.3838>
- Fienberg, S. E., Martin, M. E., & Straf, M. L. (Eds.). (1985). *Sharing research data*. National Academy Press. <https://doi.org/10.17226/2033>
- Fisher, G. (2006). The autonomy of models and explanation: anomalous molecular rearrangements in early twentieth-century physical organic chemistry. *Studies in History and Philosophy of Science Part A*, 37(4), 562-584. <https://doi.org/10.1016/j.shpsa.2006.09.009>
- Fiske, A. P. (2004). Four Modes of Constituting Relationships: Consubstantial Assimilation; Space, Magnitude, Time, and Force; Concrete Procedures; Abstract Symbolism. In N. Haslam (Ed.), *Relational models theory: A contemporary overview*, 61–146. Lawrence Erlbaum Associates Publishers. Recuperado em 09 de dezembro de 2021, de <https://www.proquest.com/docview/620480952>
- Flores, J. R., Brodeur, J. J., Daniels, M. G., Nicholls, N., & Turnator, E. (2015). Libraries and the research data management landscape. *The process of discovery: The CLIR*

- postdoctoral fellowship program and the future of the academy*, 82-102. Recuperado em 09 de dezembro de 2021, de <https://www.clir.org/wp-content/uploads/sites/6/RDM.pdf>
- Facilitate Open Science Training for European Research. (n.d.). *Open Science | FOSTER*. Recuperado em 27 janeiro de 2020, de <https://www.fosteropenscience.eu/foster-taxonomy/open-access-definition>
- Facilitate Open Science Training for European Research. (2018). *What is Open Science?* <https://doi.org/10.5281/zenodo.2629946>
- Facilitate Open Science Training for European Research. (2020). Recuperado em 22 janeiro de 2020, de <https://www.fosteropenscience.eu/taxonomy/term/9>
- Freitas, M. A. (2015). *Diretrizes para o depósito da produção científica em repositórios institucionais* [tese de doutorado]. Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília. Recuperado em 24 julho de 2021, de https://repositorio.unb.br/bitstream/10482/19189/1/2015_MaríliaAugustadeFreitas.pdf
- Freitas, M. A., & Leite, F. C. L. (2019). La Percepción de investigadores brasileños sobre el depósito de la producción científica en repositorios institucionales de acceso abierto. *Revista Interamericana De Bibliotecología*, 42(2), 159–173. <https://doi.org/10.17533/udea.rib.v42n2a04>
- Freitas, M. A., & Leite, F. C. L. (2019a). Atores do sistema de comunicação científica: apontamentos para discussão de suas funções. *Informação & Informação*, 24(1), 273-299. <http://dx.doi.org/10.5433/1981-8920.2019v24n1p273>
- Frické, M. (2015), Big Data and Its Epistemology. *J Assn Inf Sci Tec*, 66, 651-661. <https://doi.org/10.1002/asi.23212>
- Frigg, R., & Hartmann, S. (2020). Models in science. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Recuperado em 21 setembro de 2021, de <https://plato.stanford.edu/archives/spr2020/entries/models-science/>.
- Frigg, R., & Nguyen, J. (2020). *Modelling nature: An opinionated introduction to scientific representation*. Recuperado em 11 setembro de 2021, de <http://eprints.lse.ac.uk/103568/>
- Gaines, AM, (2015) From Concerned to Cautiously Optimistic: Assessing Faculty Perceptions and Knowledge of Open Access in a Campus-Wide Study, *Journal of Librarianship and Scholarly Communication* 3 (1). <https://doi.org/10.7710/2162-3309.1212>
- Gezelter, D. (2009). *What, exactly, is open science?* Recuperado em 7 agosto de 2021, de <https://openscience.org/what-exactly-is-open-science/>
- Giglia, E., & Swan, A. (2012). Open Access to data for a new, open science. *European journal of physical and rehabilitation medicine*, 48(4), 713–716. Recuperado em 7 agosto de 2021, de

<https://www.minervamedica.it/en/journals/europa-medicophysica/article.php?cod=R33Y2012N04A0713>

- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Atlas SA.
- Gordin M.D. (2012). The Textbook Case of a Priority Dispute: D. I. Mendeleev, Lothar Meyer, and the Periodic System. In: Biagioli M., Riskin J. (eds) *Nature Engaged*. Palgrave Studies in Cultural and Intellectual History. Palgrave Macmillan. https://doi.org/10.1057/9780230338029_4
- Gray, J. (2007). *eScience: a transformed scientific method. The Fourth Paradigm*. Recuperado em 13 novembro de 2021, de <https://www.coursehero.com/file/10424799/eScience-A-Transformed-Scientific-Method-midterm-reading/>
- Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. J., & Heber, G. (2005). Scientific data management in the coming decade. *Acm Sigmod Record*, 34(4), 34-41. Recuperado em 15 novembro de 2021, de <https://arxiv.org/ftp/cs/papers/0502/0502008.pdf>
- Gray, J., Szalay, A. S., Thakar, A. R., Stoughton, C., & vandenBerg, J. (2002). Online scientific data curation, publication, and archiving. In A. S. Szalay (Ed.), *Virtual Observatories 4846*, 103–107. International Society for Optics and Photonics. Recuperado em 13 novembro de 2021, de <https://www.microsoft.com/en-us/research/publication/online-scientific-data-curation-publication-and-archiving/>
- Grimm, J. (2008). *Wessex Archaeology Metric Archive Project (WAMAP)*. [data-set]. Archaeology Data Service [distribuidor]. <https://doi.org/10.5284/1000043>
- Harley, D., Earl-Novell, S., Arter, J., Lawrence, S., King, C.J. (2010). The influence of academic values on scholarly publication and communication practices. *Journal of Electronic Publishing*; 10(2). <http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0010.204>
- Harnard, S. (2005). To cognize is to categorize: cognition is categorization. In: Cohen, H., Lefebvre, C. (Eds), *Handbook of categorization in cognitive science*, 20-45. Elsevier Academic Press. <https://doi.org/10.1016/B978-0-08-101107-2.00002-6>
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Hajjem, C., & Hilf, E. R. (2008). The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review*, 34(1), 36–40. <https://doi.org/10.1080/00987913.2008.10765150>
- Hartmann, S., & Frigg, R. (2005). Scientific Models. In Sarkar Sahotra (ed.), *The Philosophy of Science: An Encyclopedia*, 2. Routledge. Recuperado em 24 outubro de 2021, de http://fitelson.org/probability/sarkar_philosophy_of_science_encyclopedia.pdf
- Harvey, R. (2010). *Digital Curation: A How-To-Do-It Manual*. Neal.
- Hedstrom, M., & Myers, J. (2018). *SEAD: Finding the long tail lost in big data*. Recuperado em 21 novembro de 2021, de <https://wiki.ncsa.illinois.edu/download/attachments/33001542/NDS%20White%20P>

- aper-%20The%20Long%20Tail%20Lost%20in%20Big%20Data-2.pdf?version=1&modificationDate=1402411325000&api=v2
- Heidorn, B. P. (2008). Shedding light on the Dark Data in the Long Tail. *Library Trends*, 57(2), 280-299. Recuperado em 12 agosto de 2021, de <https://core.ac.uk/download/pdf/4819969.pdf>
- Henning, P. C., Ribeiro, C. J. S., da Silva Santos, L. O. B., & dos Santos, P. X. (2019). GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. *Em Questão*, 25(2), 389-412. <http://dx.doi.org/10.19132/1808-5245252.389-412>
- Hess, C., & Ostrom, E. (Eds.). (2007). *Understanding Knowledge as a Commons: From Theory to Practice*. The MIT Press. Recuperado em 10 agosto de 2021, de <http://www.jstor.org/stable/j.ctt5hhdf6>
- Hesse, M. (1965). Models and Analogies in Science. *British Journal for the Philosophy of Science* 16 (62):161-163.
- Hey A.J.G., Tansley D.S.W., & Tolle K. M., (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. *in Proceedings of the IEEE*, 99(8) 1334-1337. Recuperado em 12 agosto de 2021, de <http://fourthparadigm.org/>
- Hey, A J G & Trefethen, A E (2003) The Data Deluge: An e-Science Perspective. In, Berman, F, Fox, G C and Hey, A J G (eds.) *Grid Computing - Making the Global Infrastructure a Reality*, 809-824. <https://doi.org/10.1002/0470867167.ch36>
- Higgins, S. (2008). Draft DCC curation lifecycle model. *International Journal of Digital Curation*, 2(2). <https://doi.org/10.2218/ijdc.v3i1.48>
- Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights*, 32(1), 18. <http://doi.org/10.1629/uksg.468>
- Humphrey, C. (2006). *e-Science and the Life Cycle of Research*. <https://doi.org/10.7939/R3NR4V>
- Inter-university Consortium for Political and Social Research. (2020). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle*, 6th ed., Inter-university Consortium for Political and Social Research (ICPSR), Institute for Social Research University of Michigan. <https://dx.doi.org/10.7302/3705>
- Ilyenkov, E. V. (1982). *The dialectics of the abstract and the concrete in Marx's Capital*. Progress.
- Ilyenkov, E. V. (2008). *Dialectical logic; Essays on its history and theory*. Aakar Books.
- Institute of Medicine. (2013). Sharing clinical research data: *Workshop summary*. The National Academies Press. <https://doi.org/10.17226/18267>.
- Interagency Working Group on Digital Data. (2009). Harnessing the power of digital data for science and society. In *Report of the IWGDD to the Committee on Science of the*

- National Science and Technology Council*. Recuperado em 09 de dezembro de 2021, de https://www.nitrd.gov/pubs/Report_on_Digital_Data_2009.pdf
- Jacobs, J. & Humphrey, C. (2004). Preserving research data. *Communications of the ACM*, 47(9), 27-29. <https://doi.org/10.1145/1015864.1015881>
- Joint Information Systems Committee. (2021). *Research data management toolkit*. Jisc. Recuperado em 29 de dezembro de 2021, de <https://www.jisc.ac.uk/full-guide/rdm-toolkit>
- Joint Information Systems Committee. (2011). Data centres: their use, value and impact. *Research Information Network*. Jisc. Recuperado em 03 de dezembro de 2021, de https://www.webarchive.org.uk/wayback/archive/20140615234015mp_/http://www.jisc.ac.uk/media/documents/publications/general/2011/datacentres.pdf
- Joint, N. (2006), Institutional repositories, self-archiving and the role of the library, *Library Review*, 55(2), 81-84. <https://doi.org/10.1108/00242530610649576>
- Jones, K. S., Walker, S., & Robertson, S. E. (2000). *A probabilistic model of information retrieval: development and comparative experiments*. Recuperado em 24 de dezembro de 2021, de <http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.134.6108>
- Jones, S. (2011). *Research data policy briefing*. DCC. Recuperado em 12 Março 2021, de https://www.dcc.ac.uk/sites/default/files/documents/resource/policy/DCC_policy_briefing_2011.pdf
- Jones, S., Pryor, G., & Whyte, A. (2013). How to Develop Research Data Management Services - a guide for HEIs. *DCC How-to Guides*. Digital Curation Centre. Recuperado em 12 Março 2021, de <http://www.dcc.ac.uk/resources/how-guides>.
- Jorge, V., & Albagli, S. (2017). Pierre Bourdieu e a produção do conhecimento científico: reflexões sobre uma ciência aberta. Em: Mareleto, R.; Pimenta, R. (Eds.). *Pierre Bourdieu e a produção social da cultura, do conhecimento e da informação*. Garamond, 207–228. Recuperado em 12 Março 2021, de https://edisciplinas.usp.br/pluginfile.php/4952291/mod_resource/content/1/pierre_bourdieu_ebook.pdf
- Justi, R. (2006). La enseñanza de ciencias basada en la elaboración de modelos. *Enseñanza de Las Ciencias: Revista de Investigación y Experiencias Didácticas*, 24(2), 173-184. <http://hdl.handle.net/11162/22332>
- Kaptelinin, V. (2005). The object of activity: Making sense of the sense-maker. *Mind, culture, and activity*, 12(1), 4-18. https://doi.org/10.1207/s15327884mca1201_2
- Karasti, T. H., Millerand, F., Hine, C. M., & Bowker, G. C. (2016). Knowledge infrastructures: Part I. *Science & Technology Studies*, 29(1), 2-12. <https://doi.org/10.23987/sts.55406>
- Katz, D. S., Allen, G., Barba, L. A., Berg, D. R., Bik, H., Boettiger, C., Borgman, C. L., Brown, C. T., Buck, S., Burd, R., de Waard, A., Eve, M. P., Granger, B. E., Greenberg, J., Howe, A., Howe, B., Khanna, M., Killeen, T. L., Mayernik, M.,

- McKiernan, E., ... Walsh, J. (2018). The principles of tomorrow's university. *F1000Research*, 7, 1926. <https://doi.org/10.12688/f1000research.17425.1>
- Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.
- Kenney, A.R., & McGovern, N.Y. (2005). *The three-legged stool: Institutional Response to Digital Preservation* [Slides de PowerPoint]. Recuperado em 12 novembro 2020, de http://www.library.cornell.edu/iris/dpo/docs/Cuba-ark-nym_final.ppt
- Kim, J. (2013). Data sharing and its implications for academic libraries, *New Library World*, 114(11/12), 494-506. <https://doi.org/10.1108/NLW-06-2013-0051>
- Kim, Y., & Stanton, J.M. (2016). Institutional and Individual Factors Affecting Scientists Data Sharing Behaviors: A Multilevel Analysis. *J Assn Inf Sci Tec*, 67, 776-799. <https://doi.org/10.1002/asi.23424>
- Kindling, M. (2013). Doctoral theses' research data and metadata documentation. In *ETD 2013 Hong Kong 16th International Symposium on Electronic Theses and Dissertations*, 25. Recuperado em 12 de dezembro de 2021, de <https://lib.hku.hk/etd2013/presentation/Maxi-ETD-20130925.pdf>
- Knorr-Cetina, K., & Reichmann, W. (2015). Epistemic Cultures. *International Encyclopedia of the Social & Behavioral Sciences*, 873–880. <https://doi.org/10.1016/b978-0-08-097086-8.10454-4>
- Kouper, I. (2016). Professional participation in digital curation. *Library & Information Science Research*, 38(3), 212-223. <https://doi.org/10.1016/j.lisr.2016.08.009>
- Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual review of information science and technology*, 45(1), 247-294. <https://doi.org/10.1002/aris.2011.1440450113>
- Kozlov M. (2022). NIH issues a seismic mandate: share data publicly. *Nature*, 602(7898), 558–559. <https://doi.org/10.1038/d41586-022-00402-1>
- Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning*, 3(6), 643-654. <https://doi.org/10.1504/IJTEL.2011.045454>
- Kuhn, T. S. (1996). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press. <https://doi.org/10.7208/chicago/9780226458106.001.0001>
- Kuramoto, H. (2006). Informação científica: proposta de um novo modelo para o Brasil. *Ciência da informação*, 35, 91-102. <https://doi.org/10.1590/S0100-19652006000200010>
- Kuutti, K. (1996). Activity theory as a potential framework for human-computer interaction research. Context and consciousness: *Activity theory and human-computer interaction*, 1744. Recuperado em 15 novembro 2021, de http://www.gerrystahl.net/teaching/spring09/782readings/Kuutti_AT.pdf
- Lakatos, E. M. & Marconi, M. A. (1991). *Metodologia científica*. 2. ed. Atlas.

- Lämmerhirt, D. (2016). Disciplinary Differences in Opening Research Data. *PASTEUR4OA Briefing Paper*. <http://dx.doi.org/10.2139/ssrn.3322652>
- Latour, B. (1988). *Mixing humans and nonhumans together: The sociology of a door-closer*. *Social Problems*, 35(3), 298-310. Recuperado em 10 julho 2021, de https://nissenbaum.tech.cornell.edu/papers/Latour_Mixing.pdf
- Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. Cambridge University Press.
- Lavoie, B. (2012). Sustainable research data. In G. Pryor (Ed.), *Managing Research Data*, 67-82. Facet. <https://doi.org/10.29085/9781856048910.005>
- Law, J. (1987). The structure of sociotechnical engineering: A review of the new sociology of technology. *The Sociological Review*, 35(2), 404-425. <https://doi.org/10.1111/j.1467-954X.1987.tb0001>
- Lee, C. & Tibbo, H. (2007). Digital Curation and Trusted Repositories: Steps Toward Success. *Journal of Digital Information*, 8(2). Recuperado em 09 de dezembro de 2021, de <https://journals.tdl.org/jodi/index.php/jodi/article/view/229/183>.
- Leite, F. & Costa, S. (2016). Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto. *Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información*. 30. 41-71. <https://repositorio.unb.br/handle/10482/9753>
- Leont'ev, A. N. (1981). *Problems of the Development of the Mind: A. N. Leontyev*. Progress.
- Lewis, M. (2010). Libraries and the management of research data. In S. McKnight (Ed.), *Envisioning Future Academic Library Services: Initiatives, Ideas and Challenges*, 145-168. Facet. <https://doi.org/10.29085/9781856048750.011>
- Lewis, S., Hayes, L., Newton-Wade, V., Corfield, A., Davis, R., Donohue, T., & Wilson, S. (2009) If SWORD is the Answer, What is the Question? Use of the simple webservice offering repository deposit protocol. *Program: electronic library and information systems*, 43(4), 407-418. <http://dx.doi.org/10.1108/00330330910998057>
- Lima, T. C. S., & Mito, R. C. T. (2007). Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica. *Revista Katál*. 10(esp.), 37-45. <https://doi.org/10.1590/S1414-49802007000300004>.
- Lord, P., & Macdonald, A. (2003). *Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision*. JISC Report. Recuperado em 09 de dezembro de 2021, de <https://www.digitalpreservation.gov/news/2004/e-ScienceReportFinal.pdf>
- Lord, P., Macdonald, A., Lyon, L., & Giaretta, D. (2004). From data deluge to data curation. *In Proceedings of the UK e-science All Hands meeting*, 440(2004), 371-375. Recuperado em 09 de dezembro de 2021, de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.7425&rep=rep1&type=pdf>

- Lynch, C.A. (2003). Institutional Repositories: Essential Infrastructure for Scholarship In The Digital Age. *portal: Libraries and the Academy* 3(2), 327-336. <https://doi.org/10.1353/pla.2003.0039>
- Marcial, L. H. & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61(10), 2029-2048. <https://doi.org/10.1002/asi.21339>
- Marconi, M. A. & Lakatos, E. M. (2003). *Fundamentos da Metodologia Científica*. Atlas.
- Maron, N. L., & Smith, K. K. (2008). Current models of digital scholarly communication: Results of an investigation conducted by Ithaka for the association of research libraries. *Association of Research Libraries*. Recuperado em 10 julho 2021, de <https://files.eric.ed.gov/fulltext/ED503401.pdf>
- Martins, A. B., Nolasco, B., & Silva, D. (2013). Integração de informação e interoperabilidade: o Repositório Institucional como suporte à gestão da investigação na Universidade de Aveiro. InCID: *Revista de Ciência da Informação e Documentação*, 4(2), 202-211. <https://doi.org/10.11606/issn.2178-2075.v4i2p202-211>
- Marx, K. (1976). *Capital: volume I*. Penguin Books.
- Marx, K., & Engels, F. (2015). *A ideologia alemã: Crítica da mais recente filosofia alemã em seus representantes Feuerbach, B. Bauer e Stirner, e do socialismo alemão em seus diferentes profetas*. Boitempo Editorial. Recuperado em 18 setembro 2021, de <https://books.google.com.br/books?id=FisOEAAAQBAJ>
- Massimi, M. (2019). Two Kinds of Exploratory Models. *Philosophy of Science* 86 (5):869-881. <https://doi.org/10.1086/705494>
- Masuzzo, P. & Martens, L. (2017). Do you speak open science? Resources and tips to learn the language. *PeerJ Preprints*. <https://doi.org/10.7287/peerj.preprints.2689v1>
- Mauthner, N. (2012). Are research data a common resource? *Feminists@Law*, 2(2), 1-22. Recuperado em 12 de dezembro de 2021, de <http://journals.kent.ac.uk/index.php/feministsatlaw/article/view/60/174>
- Maxwell, D. (2016). The Research Lifecycle as a Strategic Roadmap. *Journal of Library Administration*, 56(2), 111–123. <https://doi.org/10.1080/01930826.2015.1105041>
- Mayernik, M. S. (2016). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, 67(4), 973–993. <https://doi.org/10.1002/asi.23425>.
- Medyckyj-Scott, D., Stock, K., Gibb, R., Gehagan, M., Dzierzon, H., Schmidt, J., & Collins, A. (2016). *Our Land and Water National Science Challenge: A data ecosystem for land and water data to achieve the challenge mission*. Landcare Research.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press

- Michener W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. *PLoS computational biology*, 11(10), e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>
- Minayo, M. C. (2001) Ciência, técnica e arte: O desafio da Pesquisa Social. In Minayo, M. C. (Org.). *Pesquisa social: Teoria, método e criatividade*. Vozes.
- Mitroff, I. I. (1974). Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists. *American Sociological Review*, 39(4), 579–595. <https://doi.org/10.2307/2094423>
- Möller, K. (2013). Lifecycle models of data-centric systems and domains The abstract data lifecycle model. *Semantic Web*, 4(1), 67–88. Recuperado em 09 de dezembro de 2021, de [https:// http://www.semantic-web-journal.net/sites/default/files/swj125_0.pdf](https://http://www.semantic-web-journal.net/sites/default/files/swj125_0.pdf)
- Morgan, M., & Morrison, M. (Eds.). (1999). *Models as Mediators: Perspectives on Natural and Social Science* (Ideas in Context). Cambridge University Press. <https://doi.org/10.1017/CBO9780511660108>
- Morrison, M. (1999). Models as autonomous agents. *Ideas in Context*, 52, 38-65. Recuperado em 09 de dezembro de 2021, de <http://strangebeautiful.com/other-texts/morgan-morrison-mods-mediators.pdf#page=51>
- Mort, M., (2006). *The Health and Social Consequences of the 2001 Foot and Mouth Disease Epidemic in North Cumbria*. Colchester, Essex: UK Data Archive. Recuperado em 12 de dezembro de 2021, de <https://doc.ukdataservice.ac.uk/doc/5407/mrdoc/pdf/5407finalreport.pdf>
- Mulkay, M. J. (1976). Norms and ideology in science. *Social Science Information*, 15(4–5), 637–656. <https://doi.org/10.1177/053901847601500406>
- Nassi-Caló, L. (2015). Acesso Aberto na América Latina livre de periódicos predatórios. *SciELO em Perspectiva*, 13. Recuperado em 02 de fevereiro de 2021, de <https://blog.scielo.org/blog/2015/11/13/periodicos-predatorios-o-lado-obsкуро-do-acesso-aberto/>
- Niedzwiedzka, B. (2003). A proposed general model of information behaviour. *Information Research*, 9(1). Recuperado em 26 de janeiro de 2021, de <http://InformationR.net/ir/9-1/paper164.html>
- Gezelter, D. (2011). *An informal definition of open science*. Recuperado em 03 de dezembro de 2021, de <http://www.openscience.org/blog/?p=454>
- National Institutes of Health (2003). *NIH Data Sharing Policy and Implementation Guidance*. Recuperado em 03 de dezembro de 2021, de http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- Niu, J. (2009), Overcoming inadequate documentation. *Proc. Am. Soc. Info. Sci. Tech.*, 46, 1-14. <https://doi.org/10.1002/meet.2009.145046024>
- Nouvel, P. (2002). *Enquête sur le concept de modèle*. Presses universitaires de France.

- NSF (2011). *Digital Research Data Sharing and Management*. Recuperado em 03 de dezembro de 2021, de www.nsf.gov/nsb/publications/2011/nsb1124.pdf
- O'Carroll, C., Hyllseth, B., Berg, R. van den, Kohl, U., Kamerlin, C. L., Brennan, N., & O'Neill, G. (Eds.). (2017). *Providing researchers with the skills and competencies they need to practise Open Science*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/121253>
- Organização para a Cooperação e Desenvolvimento Econômico. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD Publishing. <https://doi.org/10.1787/9789264034020-en-fr>.
- Organização para a Cooperação e Desenvolvimento Econômico. (2015). Making Open Science a Reality", *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing. <https://doi.org/10.1787/5jrs2f963zs1-en>.
- Organização para a Cooperação e Desenvolvimento Econômico. (2016). *Guidelines on FAIR data management in horizon 2020*. Recuperado em 09 de dezembro de 2021, de http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- Organização para a Cooperação e Desenvolvimento Econômico. (2020). *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, OECD Publishing. <https://doi.org/10.1787/947717bc-en>.
- Organização para a Cooperação e Desenvolvimento Econômico. (2021). *Recommendation of the council concerning access to research data from public funding*. Recuperado em 03 de dezembro de 2021, de <https://legalinstruments.oecd.org/en/instruments/OECDLEGAL-0347>
- Oliveira, E., Ens, R., Andrade, D., & Mussis, C. R. (2003). Análise de conteúdo e pesquisa na área de educação. *Revista Diálogo Educacional*, 4(9), 11-27. <https://doi.org/10.7213/rde.v4i9.6479>
- Open Knowledge Foundation. (n.d.). *Open Definition - Defining Open in Open Data, Open Content and Open Knowledge*. opendefinition.org. Recuperado em 03 de dezembro de 2021, de <http://opendefinition.org/>
- Open Access Infrastructure for Research in Europe. (2017). *OpenAIRE Research Data Management Briefing Paper Understanding Research Data Management*. Recuperado em 16 de dezembro de 2021, de <https://www.openaire.eu/briefpaper-rdm-infonoads/view-document>
- Open Access Infrastructure for Research in Europe (2021). *Open Science overview in Europe*. Recuperado em 16 de dezembro de 2021, de <https://www.openaire.eu/os-eu-countries>.
- Palmer, C. L., Weber, N. M., & Cragin, M. H. (2011). The analytic potential of scientific data: Understanding re-use value. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-10. <https://doi.org/10.1002/meet.2011.14504801174>

- Pampel H., Vierkant P, Scholze F, Bertelmann R, Kindling M, Klump J, et al. (2013) Making Research Data Repositories Visible: The re3data.org Registry. *PLoS ONE* 8(11): e78080. <https://doi.org/10.1371/journal.pone.0078080>
- Pampel H., Dallmeier-Tiessen S. (2014). Open Research Data: From Vision to Practice. In: Bartling S., Friesike S. (eds) *Opening Science*. Springer, Cham. https://doi.org/10.1007/978-3-319-00026-8_14
- Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16, 8. <http://doi.org/10.5334/dsj-2017-008>
- Pelaprat, E., & Cole, M. (2011). Minding the gap: Imagination, creativity and human cognition. *Integrative Psychological and Behavioral Science*, 45(4), 397-418. <https://doi.org/10.1007/s12124-011-9176-5>
- Pepe, A., Mayernik, M.S., Borgman, C.L., & Sompel, H.V. (2010). From artifacts to aggregations: Modeling scientific life cycles on the semantic Web. *J. Assoc. Inf. Sci. Technol.*, 61, 567-582. <https://doi.org/10.1002/asi.21263>
- Pietrocola, M. (1999). Construção e realidade: o realismo científico de Mário Bunge e o ensino de ciências através de modelos. *Investigações em ensino de Ciências*, 4(3), 213-227. Recuperado em 23 de dezembro de 2021, de https://edisciplinas.usp.br/pluginfile.php/6725/mod_resource/content/1/Texto_05_-_Construcao_e_realidade.pdf
- Pigott, D., Hobbs, V., & Gammack, J.G. (2001). An Approach to Managing Repurposing of Digitised Knowledge Assets. *Australasian. J. Inf. Syst.*, 9(1). <https://doi.org/10.3127/ajis.v9i1.217>
- Pinfield, S., Cox, A.M. & Smith, J. (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0114734>
- Piwowar H., Priem J., Larivière V., Alperin J.P., Matthias L, Norlander B., Farley A., West J., & Haustein S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ* 6. <https://doi.org/10.7717/peerj.4375>
- Piwowar H.A., Day R.S., & Fridsma D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3). <https://doi.org/10.1371/journal.pone.0000308>
- Polanyi, M. (2000). The Republic of Science: Its Political and Economic Theory. *Minerva: A Review of Science, Learning & Policy*, 38(1), 1–21. <https://doi.org/10.1023/A:1026591624255>
- Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015). Fostering open science to research using a taxonomy and an eLearning portal. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business (i-KNOW '15)*. Association for Computing Machinery, Article 11, 1–8. <https://doi.org/10.1145/2809563.2809571>

- Querol, M. A. P., Cassandre, M. P., & Bulgacov, Y. L. M. (2014). Teoria da Atividade: contribuições conceituais e metodológicas para o estudo da aprendizagem organizacional. *Gestão & Produção*, 21(2), 405–416. <https://doi.org/10.1590/0104-530x351>
- Querol, M. A. P., Jackson Filho, J. M., & Cassandre, M. P. (2011). Change Laboratory: uma proposta metodológica para pesquisa e desenvolvimento da Aprendizagem Organizacional. *Administração: ensino e pesquisa*, 12(4), 609-640. <https://doi.org/10.13058/raep.2011.v12n4.143>
- Quivy, R., & Campenhoudt, L.V. (1988). *Manual de investigação em ciências sociais*. Gradiva.
- Ram, K. (2013). Git can facilitate greater reproducibility and increased transparency in science. *Source code for biology and medicine*, 8(1):7. <https://doi.org/10.1186/1751-0473-8-7>
- Ramachandran, R., Bugbee, K., & Murphy, K. (2021). From open data to open science. *Earth and Space Science*, 8. <https://doi.org/10.1029/2020EA001562>
- Raymond, E. (1999). The cathedral and the bazaar. *Know Techn* 12, 23–49. <https://doi.org/10.1007/s12130-999-1026-0>.
- Reilly, S., Schallier, W., Schrimpf, S., Smit, E., & Wilkinson, M. (2011). *Report on integration of data and publications*. Recuperado em 03 de dezembro de 2021, de <http://libereurope.eu/wp-content/uploads/ODE-ReportOnIntegrationOfDataAndPublication.pdf>.
- Relatório, UC. (2019). *Investigação na Universidade de Coimbra Relatório da Comissão de Investigação do Conselho Geral*. Recuperado em 14 de Junho de 2022, de https://www.uc.pt/governo/cons_geral/docs_aprovados_eventos/doc_CI
- Roth, W. M. (2004). Introduction: " activity theory and education: An introduction". *Mind, Culture, and Activity*, 11(1), 1-8. https://doi.org/10.1207/s15327884mca1101_1
- Roth, W. M., & Lee, Y. J. (2007). Vygotsky's neglected legacy: Cultural-historical activity theory. *Review of educational research*, 77(2), 186-232. <https://doi.org/10.3102/0034654306298273>
- Rowley, J., Johnson, F., Sbaffi, L., Frass, W., & Devine, E. (2017). Academics' behaviors and attitudes towards open access publishing in scholarly journals. *Journal of the Association for Information Science and Technology*, 68(5), 1201-1211. <https://doi.org/10.1002/asi.23710>
- Royal Society (2012). *Science as an open enterprise*. London: The Royal Society Science Policy Centre. Recuperado em 03 de dezembro de 2021, de https://royalsociety.org/~media/royal_society_content/policy/projects/sape/2012-06-20-saoc.pdf
- Salmi, J. (2015). Study on open science: Impact, implications and policy options. *European Commission Report*. Recuperado em 13 de outubro de 2021, de <https://op.europa.eu/s/u4IH>

- Sannino, A. (2005). Cultural-historical and discursive tools for analyzing critical conflicts in students' development. In K. Yamazumi, Y. Engström, & H. Daniels (Eds.), *New learning challenges: going beyond the industrial age system of school and work*, 165-195. Kansai University Press.
- Sannino, A. (2011). Activity theory as an activist and interventionist theory. *Theory & Psychology*, 21(5), 571-597. <https://doi.org/10.1177/0959354311417485>
- Sannino, A., & Engeström, Y. (2018). Cultural-historical activity theory: Founding insights and new challenges. *Cultural-historical psychology*. <http://hdl.handle.net/10138/299511>
- Sant'Ana, R. C. G. (2013). *Ciclo de vida dos dados e o papel da Ciência da Informação*. Recuperado em 09 de dezembro de 2021, de <http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>
- Sant'Ana, R. C. G. (2016). Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. *Informação & Informação*, 21(2), 116-142. <http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>
- Sayão, L. F. (2001). Modelos teóricos em ciência da informação-abstração e método científico. *Ciência da informação*, 30, 82-91. <http://dx.doi.org/10.18225/ci.inf..v30i1.941>
- Sayão, L. F., & Sales, L. F. (2014). Dados abertos de pesquisa: ampliando o conceito de acesso livre. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, 8(2). <http://dx.doi.org/10.3395/reciis.v8i2.611>
- Sayão, L. F., & Sales, L. F. (2015). *Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores*. <http://carpedien.ien.gov.br/handle/ien/1624>
- Sayão, L. F., & Sales, L. F. (2016). Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, 21(2), 90-115. <http://dx.doi.org/10.5433/1981-8920.2016v21n2p90>
- Sayão, L. F., & Sales, L. F. (2019). A grande e a pequena ciência: análise das diferenças na gestão de dados de pesquisa. *Informação & Sociedade-Estudos*, 29, 151-170. <http://hdl.handle.net/20.500.11959/brapci/147958>
- Sayão, L. F., & Sales, L. F. (2020). AFINAL, O QUE É DADO DE PESQUISA? *BIBLOS*, 34(2). Recuperado em 15 de novembro de 2021, de <https://periodicos.furg.br/biblos/article/view/11875>
- Schelig, K., & Friesike, S. (2014). Putting open science into practice: A social dilemma? *First Monday*, 19(9). <https://doi.org/10.5210/fm.v19i9.5381>
- Scribner, S. (1985). Vygotsky's uses of history. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives*, 119-145. Cambridge University Press.

- Severin, A., Egger, M., Eve, M. P., & Hürlimann, D. (2018). Discipline-specific open access publishing practices and barriers to change: an evidence-based review. *F1000Research*, 7, 1925. <https://doi.org/10.12688/f1000research.17328.2>
- Shearer, K. (2015). *Comprehensive Brief on Research Data Management Policies*. Recuperado em 15 de novembro de 2021, de <http://docplayer.net/17594465-Comprehensivebrief-on-research-data-management-policies.html>
- Shepanek, R (2011). Harnessing the power of digital data: taking the next step. Co-sponsored by the Environmental Protection Agency (EPA), CENDI (The Federal STI Managers Group), and the Interagency Working Group on Digital Data (IWGDD). Report from the workshop to improve SDM. Recuperado em 05 de dezembro de 2021, de https://cfpub.epa.gov/si/si_public_file_download.cfm?p_download_id=501807&Lab=OSIM
- Slota, S. C., & Bowker, G. C. (2017). How Infrastructures Matter in Felt, U., Fouché, R., Miller, C. A., & Smith-DoerR, L. (Eds.) *The Handbook of Science and Technology Studies*. 529-554. MIT Press.
- Soehner, C., Steeves, C., & Ward, J. (2010). E-Science and Data Support Services: A Study of ARL Member Institutions. *Association of Research Libraries*. Recuperado em 15 de novembro de 2021, de <https://eric.ed.gov/?id=ED528643>
- Stetsenko, A. (2003). Alexander Luria and the cultural-historical activity theory: Pieces for the history of an outstanding collaborative project in psychology. *Mind, Culture, and Activity*. https://doi.org/10.1207/S15327884MCA1001_10
- Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). *Primer on Data Management: What you always wanted to know*. <https://doi.org/doi:10.5060/D2251G48>
- Structural Reform Group (2004). *DDI Version 3.0 Conceptual Model*. Data Documentation Initiative. Recuperado em 25 de outubro de 2021, de <https://ddialliance.org/sites/default/files/Concept-ModelWD.pdf>
- Suber, P. (2006). *Timeline of the open access movement*. Recuperado em 05 de dezembro de 2021, de <http://www.earlham.edu/~peters/fos/timeline.htm>
- Suber, P. (2012). *Open access*. MIT Press. Recuperado em 07 de dezembro de 2021, de <http://nrs.harvard.edu/urn-3:HUL.InstRepos:10752204>.
- Swan, A., & Brown, S. (2005). *Open access self-archiving: An author study. Technical Report, Joint Information Systems Committee (Jisc)*, UK FE and HE funding councils. <http://hdl.handle.net/10150/106218>
- Tammaro, A., Matusiak, K., Sposito, F. & Casarosa, V. (2019). Data Curator's Roles and Responsibilities: An International Perspective. *Libri*, 69(2), 89-104. <https://doi.org/10.1515/libri-2018-0090>
- Taylor, A. G & Joudrey, D. N. (2008). *The organization of the information*. 3 ed. Libraries Unlimited.
- Taylor, J. M. (2001). The UK eScience programme. In: *EScience London Meeting*

- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PloS one*, *10*(8). <https://doi.org/10.1371/journal.pone.0134826>
- Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., ... & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PloS one*, *15*(3). <https://doi.org/10.1371/journal.pone.0229003>
- Tenopir, C., Dalton, E., Christian, L., Jones, M., McCabe, M., Smith, M., & Fish, A. (2017). Imagining a gold open access future: Attitudes, behaviors, and funding scenarios among authors of academic scholarship. *College and Research Libraries*, *78*(6). <https://doi.org/10.5860/crl.78.6.824>
- Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2013). Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, *39*(1), 70–78. <https://doi.org/10.1177/0340035212473089>
- Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, *36*(2), 84-90.
- The Economist. (2017). *The World's Most Valuable Resource Is No Longer Oil, but Data*. Recuperado em 12 de agosto de 2021, de <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approachantitrust-rules-worlds-most-valuable-resource>.
- The Royal Society. (2011). Science as an open enterprise. London: *The Royal Society Science Policy Centre*. Recuperado em 09 de dezembro de 2021, de https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2_012-06-20-SAOE.pdf.
- Thessen, A., & Patterson, D. (2011). *Data issues in the life sciences*. *ZooKeys*, *150*, 15–51. <https://doi.org/10.3897/zookeys.150.1766>
- Tibbo, H. R., 2015. Digital Curation Education and Training: From Digitization to Graduate Curricula to MOOCs. *International Journal of Digital Curation*, *10*(1), 144-153. Recuperado em 09 de dezembro de 2021, de https://pdfs.semanticscholar.org/5ad2/1fa42f4f9c77e62aa8659ab87291e51c0c8c.pdf?_ga=2.245022341.1699300512.1641596083-697588487.1637693846
- Tolman, C. (1981). The metaphysic of relations in Klaus Riegel's 'dialectics' of human development. *Human development*, *24*(1), 33-51. <https://doi.org/10.1159/000272623>
- Tzanova, S. (2020). Changes in academic libraries in the era of Open Science. *Education for Information*, *36*(3), 281-299. <https://doi.org/10.3233/EFI-190259>
- UK Data Archive. (2021). *About*. Recuperado em 09 de dezembro de 2021, de <https://www.data-archive.ac.uk/about/>
- UK Data Service. (2021a). *About - UK Data Service*. Recuperado em 09 de dezembro de 2021, de <https://www.dataservice.ac.uk/about>

- UK Data Service. (2021b). *About - The importance of managing and sharing data*. Recuperado em 09 de dezembro de 2021, de <https://ukdataservice.ac.uk/learning-hub/research-data-management/>
- United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on Open Science*. Recuperado em 22 de dezembro de 2021, <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>
- United Nations Educational, Scientific and Cultural Organization. (2021b). *Open access to facilitate research and information on COVID-19. UNESCO*. Recuperado em 12 de dezembro de 2021, de <https://en.unesco.org/covid19/communicationinformationresponse/opensolution>.
- Universidade de Coimbra. (2022). *Dados Fundamentais*. Recuperado em 25 de Janeiro de 2022, de <https://www.uc.pt/dados>
- University of Virginia Library. (2022). *LibGuides: Research Data Management: Research Data Lifecycle*. Recuperado em 4 de Janeiro de 2022, de <https://guides.lib.virginia.edu/c.php?g=515290&p=3522215>.
- U.S. National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (US NSF-NSB-05-40). National Science Foundation. Recuperado em 13 de setembro de 2021, de <https://www.nsf.gov/pubs/2005/nsb0540/>
- Valles, M. S. (2007). *Entrevistas cualitativas*. Centro de Investigaciones Sociológicas.
- Van den Eynden, V. (2013). Data Life Cycle and Data Management Planning. *Looking after and managing your research data*. Recuperado em 09 de dezembro de 2021, de <https://dam.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf>
- Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). *Managing and sharing data; a best practice guide for researchers*. UK Data Archive. Recuperado em 09 de dezembro de 2021, de <https://dam.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf>
- Veiga, V. (2017). *Percepção dos pesquisadores portugueses e brasileiros da área de Neurociências quanto ao compartilhamento de artigos científicos e dados de pesquisa no acesso aberto verde: custos, benefícios e fatores contextuais*. [Tese de Doutorado]. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz. <https://www.arca.fiocruz.br/handle/icict/26842>
- Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428–436. <https://doi.org/https://doi.org/10.1016/j.jbusres.2017.12.043>
- Virkkunen, J., & Newnham, D. S. (2013). A Change Laboratory in the Central Surgical Unit of Oulu University Hospital. In *The Change Laboratory*, 165-185. Sense Publishers. https://doi.org/10.1007/978-94-6209-326-3_10
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>

- Vygotsky, L. S. (1986). *Thought and language*. MIT Press.
- Wallis J.C., Mayernik M.S., Pepe A., & Borgman C.L. (2008). An Exploration of the Life Cycle of eScience Collaboratory Data. *UCLA: Center for Embedded Network Sensing*. <http://hdl.handle.net/2142/15122>
- Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE*, 8(7). <https://doi.org/10.1371/journal.pone.0067332>
- Watson, M. (2015). When will “open science” become simply “science”? *Genome Biology*, 16, 101. <https://doi.org/10.1186/s13059-015-0669-2>
- Weber, T. & Kranzlmüller, D. (2019). Methods to Evaluate Lifecycle Models for Research Data Management. *Bibliothek Forschung und Praxis*, 43(1), 75-81. <https://doi.org/10.1515/bfp-2019-2016>
- Weidner, A. J., & Alemneh, D. G. (2013). Workflow tools for digital curation. *Code4Lib Journal*, 20. Recuperado em 09 de dezembro de 2021, de <https://journal.code4lib.org/articles/8419>
- Whyte, A., & Tedds, J. (2011). Making the Case for Research Data Management. *DCC Briefing Papers*. Digital Curation Centre. Recuperado em 19 de novembro de 2021, de <https://www.dcc.ac.uk/guidance/briefing-papers/making-case-rdm>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.18>
- Willinsky, J. (2005). The unacknowledged convergence of open source, open access, and open science. *First Monday*, 10(8). <https://doi.org/10.5210/fm.v10i8.1265>
- Winget, M., Frick, C., McDonough, J., Renear, A. & Lowood, H., (2009). Digital Curation of Humanistic, Multimedia Materials: Lessons Learned and Future Directions. *Chapel Hill*, 42-43. Recuperado em 09 de dezembro de 2021, de https://www.academia.edu/2860621/Digital_Curation_of_Humanistic_Multimedia_Materials_Lessons_Learned_and_Future_Directions
- Wyborn, L., Lehnert, K., (2014). Exploiting the long tail of scientific data: Making small data BIG. In: *Eresearch Australasia Conference*. Anais... Melbourne, Australia, Recuperado em 19 de agosto de 2021, de https://eresearchau.files.wordpress.com/2016/03/eresau2016_paper_88.pdf.
- Xia, J. (2010), A longitudinal study of scholars attitudes and behaviors toward open-access journal publishing. *J. Am. Soc. Inf. Sci.*, 61. 615-624. <https://doi.org/10.1002/asi.21283>
- Yin, R. K. (2001). *Estudo de caso: Planejamento e métodos* (2a. ed.). Bookman. <https://ria.ufrn.br/jspui/handle/123456789/943>

- Yiotis, K. (2005). The Open Access Initiative: A New Paradigm for Scholarly Communications. *Information Technology & Libraries*, 24(4), 157–162. <https://doi.org/10.6017/ital.v24i4.3378>
- Zhu, Y. (2020). Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, 46(1), 41–52. <https://doi.org/10.1177/0165551518823174>
- Zimmerman, A. S. (2008). New Knowledge from Old Data The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology & Human Values*, 33(5), 631-652. <https://doi.org/10.1177/0162243907306704>
- Zinchenko, V.P. 1985 "Vygotsky's Ideas About Units for the Analysis of Mind." In Wertsch, J.V. (Ed.), *Culture, Communication, and Cognition: Vygotskian Perspectives*. Cambridge University Press, 94-118.
- Zuckerman, G. A. (2012). Developmental education: A genetic modeling experiment. *Journal of Russian and East European Psychology*, 49(6), 45–63. <https://doi.org/10.2753/RPO1061-0405490603>

APÊNDICES

Apêndice A – Convite para entrevistas

Prezado(a) investigador(a),

As Unidades de Investigação e Desenvolvimento (UI&D) têm um necessário e importante envolvimento com a gestão de dados de investigação (Research Data Management – RDM), cada vez mais crescente à medida que os grupos e agências financiadoras da ciência, nacionais e internacionais, exigem que as investigações sejam acompanhadas por um plano de gestão de dados, e que estes dados sejam partilhados em acesso aberto.

Neste contexto, torna-se fundamental uma gestão adequada e articulada dos dados de investigação, pois além de cumprir os requisitos das agências, podem minimizar e organizar as atividades atribuídas aos investigadores, dirimir as perdas de recursos informacionais, económicos, humanos, educacionais, entre outros.

Através da indicação do Instituto de Investigação Interdisciplinar, venho por meio deste, solicitar a vossa colaboração no trabalho de interesse da Universidade de Coimbra (UC) e das UI&D relativo à gestão de dados de investigação. O trabalho visa apresentar um modelo para a gestão de dados aplicado ao contexto da universidade, juntamente com recomendações para desenvolvimento de práticas articuladas para a RDM, observados a partir dos centros e projetos da UC.

Este primeiro contacto tem o intuito de convidá-lo(a) para participar de uma entrevista exploratória, que proporcione a identificação de práticas de gestão de dados nos projetos da unidade. Envio em anexo um documento norteador para tal propósito.

Ressalto, ainda, que o trabalho está inserido na tese de doutoramento na área da Ciência da Informação da faculdade de letras da UC, sob o referencial teórico da Ciência Aberta e orientação da professora Doutora Maria Manuel Borges e do professor Doutor Licínio Roque.

Certo de que ainda há questões e detalhes a serem elucidados e, caso haja interesse após esse primeiro contacto, estou à disposição juntamente com meus orientadores para esclarecimentos decorrentes e estabelecimento de uma data para uma primeira interlocução, seja de modo presencial ou online.

Cordialmente,

Wendell Lima

*Doutorando em Ciência da Informação
Universidade de Coimbra, Portugal.*

Apêndice B – Roteiro das entrevistas

BLOCO I – AMBIENTE E CARACTERÍSTICAS DOS DADOS NA INVESTIGAÇÃO

1. Qual o propósito do projeto de investigação?
2. Em qual domínio e finalidade os dados e a investigação estão inseridos?
3. Como é conduzido a investigação em relação a capacidade da unidade de investigação?
4. Quais são os tipos de dados e formatos produzidos e administrados?
5. Sabe informar o volume de dados geridos na investigação?
6. Como os dados administrados podem ser úteis para outros domínios do conhecimento e para a sociedade, quando pensados em utilidades diferentes do propósito original?
7. Na percepção de valor, os dados geridos são resultado de quanto tempo de trabalho?
8. Os dados apresentam informações reservadas, sensíveis e/ou pessoais?

BLOCO II – RECURSOS PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO

9. No projeto, há alguém responsável pela gestão dos dados de investigação?
10. Quem são as pessoas que estão envolvidas nas atividades para a gestão dos dados?
11. Na equipa existe alguém com conhecimentos específicos sobre a gestão de dados?
12. Existe algum fluxo de trabalho estabelecido?
13. De um modo geral, como é realizado o processo de organização dos dados?
14. Há o costume de trabalhar de forma colaborativa com instituições internas ou externas à universidade ou com profissionais da área de dados ou informação, tais como informáticos e bibliotecários? Se houver, pode descrever como ocorre?
15. Em sua percepção, quais perfis poderiam colaborar com a gestão dos dados na investigação?
16. Existem serviços ou ferramentas utilizadas para gerir os dados?
17. Como é realizado o armazenamento dos dados?
18. Há cópias de segurança dos dados?
19. Quais os meios em que os dados são partilhados entre os integrantes do projeto?
20. Há dificuldades específicas no uso de ferramentas, serviços ou infraestrutura por parte dos investigadores ou de quem realiza a gestão e a partilha dos dados?
21. Já recebeu ou recebe algum suporte em relação as ferramentas e serviços?

22. De um modo geral, utiliza-se alguma política, orientação, normas, princípios ou recomendações para a gestão e a partilha dos dados, sejam documentos internos, nacionais ou internacionais?
23. O projeto faz uso de algum plano de gestão de dados?
24. Há alguma norma ou orientação técnica quanto as questões legais, direitos autorais e de propriedade intelectual em relação aos dados de investigação?
25. Os dados ou o conjunto de dados são acompanhados de alguma informação ou descrição (metadados) para que possam ser contextualizados?

BLOCO III – HÁBITOS E PERCEPÇÕES PARA A PARTILHA EM ACESSO ABERTO

26. Os dados são disponibilizado em vias de acesso aberto? Se sim, como acontece?
27. Há dificuldades para este tipo de partilha entre os investigadores?
28. Em sua percepção, existe interesse em partilhar os dados em acesso aberto?
29. Quais são as preocupações existentes em relação a partilha e ao acesso aberto?
30. Em sua percepção, quais são as motivações e as justificativas que colaboram para a realização da partilha dos dados em acesso aberto?