



Paulo Fernando Pereira da Silva

As Políticas de Open Data em Portugal: análise da sua implementação e impacto

Dissertação de Mestrado em Ciência da Informação, orientada pela Doutora
Maria Manuel Borges e coorientada pela Doutora Daniela de Filippo,
apresentada ao Departamento de Filosofia, Comunicação e Informação da
Faculdade de Letras da Universidade de Coimbra

2017



UNIVERSIDADE DE COIMBRA

Imagem: Open Data. CC BY-NC-ND <http://www.chefuturo.it/>

Faculdade de Letras

As políticas de Open Data em Portugal:

Análise da sua implementação e impacto

Ficha Técnica:

Tipo de trabalho	Dissertação de Mestrado
Título	As políticas de Open Data em Portugal: análise da sua implementação e impacto
Autores	Paulo Fernando Pereira da Silva
Orientador	Prof. Doutora Maria Manuel Borges
Coorientador	Prof. Doutora Daniela de Filippo
Identificação do Curso	2º Ciclo em Ciência da Informação
Data da entrega	14-06-2017
Data de realização das provas	28-07-2017
Membros do Júri	
Presidente	Prof. Doutora Maria da Graça Melo Simões
Vogais	Prof. Doutora Maria Manuel Lopes Figueiredo Costa Marques Borges Prof. Doutor Andrés Pandiella Dominique
Classificação	19



*Aos meus pais, Gracinda e
Fernando.*

*À Sara, namorada,
companheira e amiga.*

AGRADECIMENTOS

O sucesso neste percurso académico deve-se também a todos os que me apoiaram, incentivaram e inspiraram, nos melhores e nos piores momentos, tornando o caminho mais fácil de percorrer. Espero que as palavras seguintes façam justiça ao sentimento de agradecimento que pretendo transmitir.

À Professora Doutora Maria Manuel Borges, a quem devo a introdução a este tema, pela total disponibilidade, pelo aconselhamento académico e profissional, espírito crítico, pela confiança depositada, pela sua amizade e pelo privilégio de poder trabalhar com alguém por quem se nutre admiração.

À Professora Doutora Daniela de Filippo pelo aconselhamento, orientação, espírito crítico e amizade que foram importantes durante o último ano.

À Sara, cujo apoio, paciência e amor foram fundamentais ao longo deste percurso.

Aos meus pais, que nunca me deixaram desistir deste sonho.

À Ana Luísa Ribeiro, colega e amiga cujo esforço me permitiu assistir a todas as aulas do mestrado.

À Cristina Caetano, colega de mestrado e amiga, pelo apoio, incentivo, partilha de ideias e espírito crítico.

Aos colegas Daniel Gonçalves, Rosa Gomes e Quele pelo companheirismo, amizade, boa energia e pelas trocas de ideias ao almoço às quintas-feiras.

A todos os professores deste ciclo académico pela mestria e paixão com que transmitiram conhecimentos e me tornaram melhor profissional da informação.

A mente que se abre a uma nova ideia, jamais volta ao seu tamanho inicial.

Albert Einstein

RESUMO

A abertura de dados científicos e de dados do setor público é garantida por políticas de Dados Abertos, que na sua constituição recomendam ou mandatam a livre disponibilização de dados. A um nível macro, a União Europeia através do seu programa de financiamento de investigação implementou um piloto de Dados Abertos, que atualmente abrange todas as áreas de investigação financiadas. O acordo de financiamento mandata a abertura de todos os dados de investigação subjacentes às publicações científicas. A União Europeia tem também publicado legislação referente à abertura de dados do setor público que é transposta para as normas jurídicas de cada estado-membro, onde Portugal não é exceção. Foram também observadas as diversas iniciativas que visam medir a abertura de dados públicos a nível europeu e mundial.

O objetivo geral do trabalho é analisar a implementação de políticas de Dados Abertos em Portugal. Dado o diferente enquadramento legislativo relativamente aos dados de investigação e aos dados do setor público, foram elencadas as políticas existentes para os diferentes tipos de dados. A abertura de dados de investigação situa-se ainda numa fase prematura, não sendo ainda visíveis resultados da implementação da Resolução do Conselho de Ministros nº21/2016. Ao nível dos dados do setor público, abertura processou-se através da transposição da Diretiva 2003/98/CE para as normas jurídicas portuguesas em 2007 e novamente em 2016. Usando o modelo de Zuiderwijk e Janssen (2014) foi elaborada uma estrutura para a avaliação do conteúdo, contexto e indicadores de desempenho da política de abertura de dados do setor público, permitindo obter uma visão geral do valor público alcançado. Foram confrontados os resultados obtidos por esta análise com os indicadores de abertura provenientes de algumas iniciativas de avaliação de impacto da abertura de dados do setor público e com a informação prestada pela AMA. O ecossistema português de Dados Governamentais Abertos encontra-se já numa fase mais avançada, revelando a existência do portal de dados nacional (dados.gov) e de diversos portais de entidades públicas e até do setor privado que disponibilizam dados de forma aberta. Esses portais foram analisados para se perceber se os conjuntos de dados cumpriam os requisitos de abertura definidos pela Open Knowledge International.

Palavras-chave: Dados Abertos, Dados Governamentais Abertos, políticas de Dados Abertos, impacto, avaliação.

ABSTRACT

The openness of scientific data and public sector data is granted by Open Data policies which recommend or mandate free availability of data. At a macro level, the European Union through its research funding program implemented an Open Data pilot that now is extended to all the funded research areas. The model agreement sets the openness of all the data underlying scientific publications. The European Union has also published legislation referring to the openness of public sector information, which is transposed to the legal system of each state member, where Portugal is no exception. We also analysed the different initiatives that aim to measure the openness of public data at European and global level.

The main goal of this academic work is to analyse the implementation of Open Data policies in Portugal. Given the different legal framework between scientific data and public sector data, we've considered and listed the policies regarding both type of data. The openness of scientific data is still in a premature phase which means that the results of the implementation of Resolução do Conselho de Ministros nº21/2016 are not visible yet. In terms of public sector data, the openness has started with the transposition of the EU Directive 2003/98/EC to the Portuguese juridical system in 2007 and again in 2016. Using Zuiderwijk e Janssen's (2014) model, it was elaborated a structure to assess the context, content and performance indicators of public sector open data policies which allows to obtain a general vision of public value achieved. The results produce by this analysis were confronted with the openness indicators generated by some public sector data impact assessment initiatives and also with the information disclosed by AMA. The Portuguese ecosystem of Open Governmental Data is in a more advanced phase, disclosing the existence of a national data portal (dados.gov), and data portals from public institutions and even from the private sector which made data available in an open manner. These data portals were analysed to understand if they fulfil all the openness requirements defined by Open Knowledge International.

Keywords: Open Data, Open Government Data, Open Data policies, impact, assessment

Sumário

AGRADECIMENTOS	v
RESUMO	ix
ABSTRACT	xi
Introdução	1
Indicações Metodológicas	6
1. Do Acesso Aberto aos Dados Abertos	8
1.1 Acesso Aberto.....	8
1.1.1 Via Dourada.....	14
1.1.2 Via Verde.....	15
1.2 A Ciência Aberta.....	17
1.3 <i>Big Data</i> ou dados massivos.....	24
1.4 Dados Abertos.....	29
1.5 Dados Abertos Governamentais.....	40
2 As políticas de Dados Abertos	53
2.1 Políticas Macro.....	70
2.2 O caso português.....	78
Conclusão	103
Referências Bibliográficas	115
LISTA DE SIGLAS	134
ÍNDICE DE FIGURAS	136
ÍNDICE DE QUADROS	137
ÍNDICE DE GRÁFICOS	138

Introdução

Uma marca distintiva da nossa época é o crescimento exponencial da produção de informação, com a proliferação de novas tecnologias de informação e comunicação que permitiram o aumento de troca de informação, de número de artigos científicos publicados e a produção massiva de dados. O acesso generalizado à internet propiciou uma alteração no paradigma do acesso e publicação de conteúdos científicos. É com o intuito de promover o livre acesso à informação científica que nasce o movimento de Acesso Aberto (AA), apoiado em três declarações de princípios de abertura, Budapeste, Bethesda e Berlim.

Segundo a Budapest Open Access Initiative (BOAI), entende-se por AA a disponibilização na Internet de literatura científica, permitindo que os utilizadores possam ler, copiar, distribuir, imprimir e pesquisar esta literatura sem barreiras legais, técnicas ou financeiras que não o próprio acesso à Internet (Budapest Open Access Initiative [BOAI], 2002, tradução livre). O AA possuiu dois modelos de abertura: a via dourada que remete para a publicação em revistas científicas de AA, e a via verde, que pressupõe o auto arquivo num repositório adequado. Subjacente a este movimento está a abertura sem qualquer tipo de discriminação de toda a literatura de investigação científica revista por pares, no caso dos postprints, como artigos, dissertações e teses, e mais recentemente, dados de investigação e livros científicos (Swan, 2012). Existem dois tipos de AA: o AA grátis, que se refere à remoção de barreiras de preço, e o AA livre, que se refere à remoção de barreiras de preço e de pelo menos algumas barreiras de permissão (Suber, 2012).

O AA enquadra-se num contexto científico mais vasto, o da Ciência Aberta (CA). O conceito de CA é definido como o esforço para disponibilizar todos os resultados da investigação científica que sejam financiados pelo setor público (Organização para a Cooperação e Desenvolvimento Económico [OECD], 2015a). Nesse sentido, a Ciência Aberta surge como um termo guarda-chuva referente à abertura de todos os resultados da investigação financiada e que assenta em vários pilares: remoção das barreiras de partilha, construção da infraestrutura para a sua concretização e como desambiguação de autores, instituições e produção científica. A CA é um meio para garantir um sistema de investigação mais eficiente, através da redução de custos no processo de coleta, transferência e reutilização de dados, do aumento da transparência e qualidade do processo de investigação, da maior rapidez na transferência de conhecimento, do aumento do impacto económico e aproximação entre ciência e cidadãos (OECD, 2015a).

A complexidade da abertura dos resultados de investigação reside no fator cultural, na implementação de um ambiente de partilha. Leonelli, Spichtinger e Prainsack

(2015:13) defendem que “in addition to the stick of compulsory mandates, carrots are therefore also needed”. As recompensas são definidas em três eixos: reconhecimento das práticas de partilha nas estruturas de incentivos, reconhecimento do papel das métricas alternativas na avaliação dos investigadores e a criação de incentivos mais significativos para envolver os investigadores na CA. Fecher e Friesike (2013) sintetizaram as mudanças operadas pela CA em cinco escolas de pensamento: ao nível da infraestrutura, da acessibilidade pública, da avaliação de impacto, do acesso gratuito e da melhoria da eficiência.

A União Europeia (EU) tem desempenhado um papel central na promoção da CA, quer através da criação de legislação adequada aos princípios de abertura que defende, quer ao nível de financiamento para o AA às publicações científicas, criação de e-infraestruturas de alojamento de recursos de informação e programas de treino e apoio. Em Portugal, apesar da já existência de e-infraestruturas científicas como o Repositório Científico de Acesso Aberto de Portugal (RCAAP) e o Serviço de Alojamento de Repositório de Dados Científicos (SARDC), apenas em 2016 se veio a definir os princípios da implementação da política nacional de CA através da Resolução do Conselho de Ministros nº21/2016. A estratégia assenta em quatro eixos: AA e Dados Abertos, infraestruturas e preservação digital, avaliação científica e responsabilidade social científica (Comissão Executiva GT-PNCA, 2016).

O crescimento exponencial da produção de dados, estimado em 40% ao ano (Manyika, et al., 2011), levou à emergência do fenómeno *Big Data* ou dados massivos, através do incremento da disponibilização de conjuntos de dados e também do reconhecimento do potencial valor da combinação e ligação de diferentes *datasets* (Ubaldi, 2013). Os dados digitais estão presentes em todos os setores económicos, existindo diversas ferramentas que coletam dados de forma massiva que depois podem ser reutilizados para desenvolver novas investigações ou aplicações¹ para benefício da sociedade. O crescimento da disponibilidade de dados originou mudanças no paradigma da investigação científica, sendo que Jim Gray considera que vivemos atualmente no 4º paradigma, caracterizado pelo uso massivo de dados na investigação científica (Lynch, 2009). Esta nova ciência assenta em três pilares: captura, curadoria e análise de dados (Bell, 2009, Royal Society, 2012). A influência do *Big Data* na sociedade manifesta-se a diversos níveis, destacando-se o desenvolvimento de projetos na área da saúde, transportes, agricultura, ambiente, indústria, ganho de competências científicas, mercado de dados e dados abertos.

Os dados que produzimos são globais, diversificados, complexos e, sobretudo, podem representar um enorme valor científico, económico, social ou cultural. Neste

¹ O termo aplicações deve ser aqui entendido de duas formas: como aplicação na inovação científica, produzindo, por exemplo, um avanço médico, e também como aplicação no sentido de software, fornecendo um serviço com o recurso à reutilização de dados.

sentido, o AA a dados torna-se um dos pilares da investigação científica do século XXI, mas também um dos pilares do desenvolvimento económico, do envolvimento dos cidadãos na ciência, através da denominada ciência cidadã, e também na auscultação da transparência governativa com a disponibilização em AA de dados governamentais.

O mundo académico tem vindo a ser pressionado para apoiar a abertura através das imposições das entidades de financiamento e do aumento de benefícios dos Dados Abertos, embora nem todos os domínios evoluam com a mesma rapidez (Hahnel & Hook, 2016). A Royal Society (2012) sublinha que os dados devem ser geridos de modo proficiente para as instituições obterem os benefícios proporcionados pela abertura de dados e entenderem a enorme mudança cultural necessária para atingir este objetivo.

Existem várias definições de Dados Abertos, apresentando algumas diferenças entre elas, embora focando o mesmo objetivo. A Royal Society (2012) entende por *Open Data* os dados que devem ser acessíveis, utilizáveis, avaliados ou certificados e inteligíveis. Contudo, “Openness is not however enough. Data must be intelligently open” (Science International, 2015:4). Segundo a *Open Definition*, os Dados Abertos englobam dois tipos diferentes de abertura: técnica e legal (Open Knowledge International [a]).

Com o objetivo de contribuir para a adoção de padrões e políticas de Dados Abertos de investigação, algumas organizações lançaram iniciativas e princípios de Open Data destinadas à recomendação de boas práticas na publicação de dados de investigação obtidos através de financiamento público. Estes princípios foram comparados de forma a obter-se uma visão geral da abertura técnica e legal recomendada.

A abertura dos dados governamentais teve a sua origem nas leis de “Freedom of Information”. Os Governos e as sociedades têm vindo de forma crescente a apoiar o acesso à informação do setor público como forma de aumentar a transparência governativa e o crescimento económico. A abertura de dados assume a preparação das entidades públicas para o processo de abertura e que o governo está disposto a efetuar mudanças consideráveis no setor público, devendo atuar como um sistema aberto que interage com o seu ambiente (Jansen, Charalabidis & Zuiderwijk, 2012).

Os Dados Governamentais Abertos (DGA) referem-se à informação coletada, produzida ou paga pelo setor público, guardada em formato que possa ser lido por máquina e tornada acessível de forma livre para reutilização para qualquer propósito, através da publicação sob uma licença aberta que permite a livre reutilização dentro do domínio público e privado (Koski, 2015; Carrara, Oudkerk, Steenbergen, Tinholt, 2016). A Open Knowledge International (OKI) (2012a) identificou as componentes chave da abertura de dados públicos: disponibilidade e acesso, reutilização e redistribuição e a participação universal. A abertura dos dados governamentais foi apoiada por várias iniciativas com o objetivo de estabelecer padrões e princípios para a

disponibilização aberta de dados, as quais foram comparadas para obter uma visão geral da abertura técnica e legal recomendada.

Para gerar valor “data must be reusable, but more importantly, actually used” (Open Data for Development [OD4D], 2016:8). O acesso e a reutilização dos dados governamentais têm um potencial significativo no aumento da eficiência e transparência do setor público, alavancando atividades de ciência cidadã com os dados governamentais, aumentando assim o seu valor económico e social (OECD, 2015a). Segundo Ubaldi (2013, tradução livre), podem ser identificadas quatro grandes fases na cadeia de valor dos Dados DGA: geração de dados, coleta, agregação e processamento de dados, distribuição e fornecimento de dados e reutilização de dados. Existem três tipos de valor associados aos DGA: valor económico, valor social e transparência pública (Ubaldi, 2013; OECD, 2015a).

O acesso a dados sem restrições de acesso deve ser garantido por uma política de Dados Abertos, pois “Open Data is political” (World Wide Web Foundation, 2016:38). Existem políticas diferenciadas para a abertura de dados governamentais e para a abertura de dados científicos, sendo que a expressão “Política de Dados Abertos” surge frequentemente associada a termos como estratégias, programas ou iniciativas de abertura de dados e ainda a políticas de gestão de dados. A contextualização das políticas não é clara, mas foca-se preferencialmente apenas nos dados provenientes do setor público. São associadas políticas de Dados Abertos de vários níveis, desde o governamental em termos de dados públicos, às políticas editoriais de abertura de dados científicos.

A nível de dados científicos salienta-se a existência de várias recomendações no sentido de balizar indicadores fundamentais para o desenvolvimento de políticas de abertura de dados científicos e a importância da gestão de dados no seio dessa abertura. A mudança cultural, a necessidade de treino dos investigadores e a abertura de dados mediante os princípios FAIR que pretendem facilitar o acesso, a interoperabilidade e a reutilização de dados são outros desafios focados. A abertura dos dados de investigação é incluída no programa de investimento Horizonte 2020 (H2020) da EU, com a criação um projeto-piloto de Dados Abertos para sete áreas, com opção de não participação, tendo sido alargado a todas as áreas científicas em projetos iniciados a partir de 2017. O artigo 29.3 do acordo modelo de financiamento requer o depósito de todos os dados e metadados associados aos resultados finais da investigação num repositório, devendo estes cumprir os princípios FAIR. A este *model grant agreement* está associada também a necessidade obrigatória de desenvolver um plano de gestão de dados que descreva os dados que serão produzidos e como e quando serão publicados. O AA a dados científicos possui um contexto legislativo diferenciado dos dados públicos, apresentando-se em Portugal numa fase ainda pouca maturada. A primeira política de Dados Abertos surge em 2014 através da Fundação

para a Ciência e Tecnologia (FCT), com o objetivo de encorajar e promover a disponibilização de dados em AA. O acesso a dados foi incluído na política científica através da Resolução do Conselho de Ministros n.º21/2016, que integrada na política nacional de CA, instaura a obrigatoriedade da abertura dos dados científicos resultantes de projetos financiados por fundos públicos.

O desenvolvimento de uma política de abertura de dados governamentais deve considerar algumas questões como a definição da sua aplicação e os objetivos, dados publicados, condições de publicação de dados, local, frequência e razão da publicação, devendo ainda ser calculado o impacto e os benefícios da abertura de dados (Carrara, et al., 2016a). Os objetivos das políticas de Dados Abertos podem ser sintetizados em três áreas principais: fomentar a inovação social e crescimento económico, melhorar a eficiência dos serviços do setor público e fortalecer a democracia e aumentar a transparência (Zuiderwijk, Gasco, Parycek & Janssen, 2014).

A EU e a Comissão Europeia têm promovido os DGA através da publicação de Diretivas Europeias (2003 e 2013) que estabeleceram regulamentos para permitir o livre acesso e reutilização dos dados produzidos pelas entidades europeias e governamentais. Esta legislação é de cariz obrigatório para todos os estados-membros. Foram também estabelecidas diversas e-infraestruturas para alojar e apoiar toda a produção de dados produzidos na EU e que indexam os dados existentes nos portais nacionais de cada país. São abordadas as várias iniciativas europeias e internacionais de avaliação de abertura de dados que permitem obter uma vista geral, ainda que desfocada, do estado de abertura de dados públicos a nível nacional e europeu.

Portugal confere o direito ao acesso à informação pública desde 1993. As políticas de abertura de dados governamentais mais recentes foram publicadas em 2007 e 2016 e assentam na transposição da Diretiva 2003/98/CE para a lei nacional. Assinala-se a existência de duas políticas municipais, a criação do portal nacional e de vários portais regionais e institucionais. Para avaliar a abertura dos dados governamentais em Portugal foi usada a estrutura elencada por Zuiderwijk e Janssen (2014). Através da análise de vários indicadores, esta estrutura fornece uma perspetiva do contexto e ambiente da política, do seu conteúdo e dos indicadores de desempenho, permitindo assim inferir o valor público alcançado. Foi também analisada a abertura de dados alojados nos portais de dados portugueses, com o recurso às características de abertura enunciadas na Open Definition.

O objetivo geral deste trabalho é identificar as políticas de Dados Abertos em Portugal, bem como a analisar a sua implementação e impacto. Neste sentido, foram delineados os seguintes objetivos específicos:

- a) Contextualizar o movimento de Dados Abertos como continuação do AA e também dentro de um modelo mais amplo, a CA, focando ainda a produção massiva de dados;
- b) Identificar princípios de abertura de dados científicos e dados públicos;
- c) Contextualizar os modelos de políticas de Dados Abertos e formas de avaliação de impacto das mesmas;
- d) Identificar a constituição e implementação de políticas de abertura de dados científicos e dados públicos a nível da EU;
- e) Identificar a constituição de políticas de abertura de dados científicos e dados públicos a nível português;
- f) Avaliação da implementação e do impacto das políticas de Dados Abertos em Portugal, em termos de dados científicos e de dados públicos.

Indicações Metodológicas

Para cumprir os objetivos estipulados, seguiu-se uma abordagem qualitativa baseada na revisão bibliográfica de artigos científicos recuperados em bases de dados especializadas disponíveis na Universidade de Coimbra, articulando a recuperação de documentos com os documentos mais citados da área da Ciência Aberta, Acesso Aberto e Dados Abertos indexados na Web of Science. Foram usados os termos de pesquisa em língua inglesa: *open science*, *open access*, *open data*, *open government data* e *scientific open data* para situar o tema, seguindo-se a pesquisa por *open data policies*, *government open data policies* e *open data impact*. A base da literatura revista para analisar o desenvolvimento e impacto das políticas de Dados Abertos em Portugal foi obtida em grande parte com o recurso ao motor de busca *Google*, derivado à especificidade da legislação, das comunicações políticas e institucionais, dos portais de dados e também à escassez de literatura científica nesse campo. Foi ainda solicitado um esclarecimento de dúvidas à Agência para a Modernização Administrativa (AMA) por correio eletrónico em termos de conteúdos e funcionalidades do portal de dados nacional, tendo as respostas do gestor de projeto André Lapa sido incluídas como contraposição a algumas afirmações da literatura ou das iniciativas de avaliação. Os Dados Abertos enquanto alvo do nosso estudo são enquadrados na dimensão da Ciência Aberta que defende a abertura de todos os resultados científicos e na dimensão de abertura da informação do setor público, que defende o acesso do cidadão à informação dos órgãos públicos e governativos.

Em termos de Dados Abertos e as suas políticas, a oferta de literatura é consideravelmente mais curta, sendo que apenas os documentos mais antigos apresentam número de citações com algum relevo. Sempre que necessário recorreu-se a trabalhos citados nas publicações em estudo. A singularidade da área e o seu rápido crescimento levou à opção por autores e artigos menos cotados mas

diretamente relacionados com o mundo dos Dados Abertos. A recolha de literatura para investigação visou identificar legislação referente a políticas de Dados Abertos, publicações de organizações internacionais, documentos oficiais relacionados com a abertura de dados e também comunicações políticas, quer para o caso português, quer ao nível macro da EU. Foi identificado o contexto social e político dos Dados Abertos, destacando-se as iniciativas com potencial impacto na abertura de dados. Foram pesquisados documentos de investigação relativos a políticas de dados abertos e formas de análise das mesmas.

A avaliação de impacto das políticas de Dados Abertos em Portugal foi feita com base na estrutura elencada por Zuiderwijk e Janssen (2014). A identificação das respostas aos 27 indicadores foi feita com o recurso à literatura, legislação e às respostas obtidas a partir da AMA.

Para determinar a quantidade de conjuntos de dados existentes em cada portal internacional analisado, quando não estava disponível a possibilidade de pesquisar por país, foi introduzido o termo Portugal na caixa de pesquisa. Deste modo, recuperaram-se os conjuntos de dados relacionados com o país, embora nem todos os *datasets* tenham autoria portuguesa. A determinação das características dos portais de dados portugueses analisados foi efetuada com base na informação inscrita no próprio sítio web. Quando não disponíveis na página principal dos catálogos de dados, procedeu-se à pesquisa nas páginas descritivas dos *datasets* ou no setor dos avisos legais.

Procurou-se aqui também definir alguns dos termos que irão ser utilizados com frequência na língua inglesa e também o seu âmbito de aplicação. Segundo Borgman (2015), dados são representações de observações, objetos ou outras entidades usadas como prova de um fenómeno para fins de investigação ou académicos.

O termo *dataset* designa um conjunto ou coleção de dados, publicados ou preservados por um agente único, estando disponível para acesso ou *download* em um ou mais formatos (World Wide Web Consortium [W3C], 2014).

Para a contextualização e avaliação de políticas foi tida em consideração a definição do Dicionário de Cambridge, que define política como “a set of ideas or a plan of what to do in particular situations that has been agreed to officially by a group of people, a business organization, a government, or a political party” (Cambridge University Press, 2017).

1. Do Acesso Aberto aos Dados Abertos

O processo de abertura do conhecimento científico começou em 2002 com o movimento do Acesso Aberto (AA), culminando na abertura da literatura científica financiada por fundo públicos. O avançar da tecnologia, a produção massiva de dados, as redes sociais científicas, a ciência cidadã, os recursos educacionais e o código aberto, em conjunto, gerou uma ciência diferente. Este conceito mais amplo de ciência, definido como Ciência Aberta, pretende abrir toda a panóplia de formatos e recursos onde constem resultados científicos, com destaque para os dados de investigação financiados publicamente. A abertura de dados processa-se também a nível público, com o Governo e várias instituições públicas empenhadas em disponibilizar de forma aberta a informação do setor público.

1.1 Acesso Aberto

O acesso ao conhecimento é a base do processo científico, motor do desenvolvimento social, económico e humano. O acesso generalizado à Internet alterou o paradigma da comunicação académica pela introdução experimental de novas formas de publicação e disseminação da ciência. Os primórdios do Acesso Aberto (AA) à informação científica remontam a iniciativas como o ArXiv², um servidor de *preprints*³, e o EPrints, um dos *softwares* usados para a constituição de repositórios.

O movimento de AA surgiu como resposta ao problema conhecido como ‘Crise dos Periódicos’. Na verdade, a escalada de preços das assinaturas de revistas científicas e as restrições colocadas pelas editoras, limitando de forma séria o acesso aos resultados de investigação, teve duas consequências: limitação severa no acesso à informação e colateral diminuição do potencial de impacto de citação. É neste contexto que em 2002 surge a declaração fundadora do AA, a Budapeste Open Access Initiative (BOAI). Segundo a BOAI, entende-se por AA a disponibilização na Internet de literatura científica, permitindo que os utilizadores possam ler, copiar, distribuir, imprimir e pesquisar esta literatura sem barreiras legais, técnicas ou financeiras que não o próprio acesso à Internet (Budapest Open Access Initiative [BOAI], 2002, tradução livre). A BOAI estabeleceu duas vias para concretizar o AA: a publicação em revistas científicas de AA, conhecida como a via dourada, e o auto arquivo em repositórios científicos, conhecido como a via verde de AA. Para atingir a abertura dos artigos, a BOAI sublinha que devem ser cumpridas duas estratégias complementares:

² O arXiv surgiu em 1991 com o intuito de alojar preprints de artigos científicos da área da física, tendo sido expandido para integrar áreas como a matemática, astronomia, ciências da computação, biologia e finança quantitativa e estatística. Disponível na WWW: <https://arxiv.org/>

³ O preprint é uma versão de um artigo científico não publicado e ainda não sujeito a revisão por pares.

- Os investigadores devem proceder ao Auto arquivo da versão completa das suas publicações;
- Necessidade de desenvolver uma nova geração de revistas científicas cujo foco seja a publicação em AA (BOAI, 2002).

Seguiram-se mais duas declarações de elevada importância para o estabelecimento do AA, a Declaração de Bethesda sobre a Publicação em Acesso Aberto e a Declaração de Berlim sobre Acesso Livre ao Conhecimento nas Ciências e Humanidades, ambas em 2003. A Declaração de Bethesda define o AA às publicações quando é autorizado o livre acesso, uso e redistribuição e também quando é fornecida a versão completa, incluindo materiais suplementares e respetiva licença de utilização, em formato digital num repositório adequado (Declaração de Bethesda sobre a Publicação em Acesso Aberto, 2003). A Declaração de Berlim impõe o cumprimento de duas condições: os autores/detentores dos direitos de autor devem autorizar, sob a forma de uma licença, qualquer tipo de cópia, reutilização, redistribuição e transmissão sujeito apenas à menção de atribuição de crédito e os autores devem depositar uma cópia completa do seu trabalho num repositório que assegure a sua preservação a longo prazo (Declaração de Berlim sobre Acesso Livre ao Conhecimento nas Ciências e Humanidades, 2003, tradução livre).

Borges (2006:92) sintetizou outros exemplos de iniciativas de AA desencadeadas a nível global, como o Open Archives Forum em 2001, a International Scholarly Communications Alliance ao nível das “bibliotecas especializadas de investigação”, listando ainda princípios editoriais dos Estados Unidos da América (EUA)⁴ e europeus⁵ e assinalando também a existência do “DAREnet, a rede dos Digital Academic Repositories”.

Em 2012, a BOAI declarou que “Nothing in the last ten years makes OA less necessary or less opportune”, renovando e definindo recomendações para os próximos 10 anos ao nível de políticas, licenciamento e reutilização, infraestrutura e sustentabilidade e por fim, sobre promoção e coordenação.

A abertura requerida por estas declarações tem por objetivo disponibilizar o AA a todos os que manifestem interesse nele, sem distinções discriminatórias. Suber (2012), um dos promotores do Movimento de Acesso Aberto à informação científica, afirma que temos de ver o conhecimento como um bem público e não como uma mercadoria. O AA deve incluir também o processamento da informação por máquinas que, através de técnicas de extração, indexação ou mistura, permitem obter novas inferências de

⁴ Os princípios foram denominados Washington D.C. Principles For Free Access to Science em 2004, tendo sido assinados “por uma série de editores não lucrativos de sociedades científicas” (Borges, 2006:92).

⁵ Borges (2006) indica que a Association of Learned and Professional Society Publishers emitiu princípios semelhantes aos princípios anunciados nos EUA.

conhecimento (Suber, 2012; Swan, 2012). Os conteúdos comumente abrangidos pelo AA são a literatura de investigação científica revista por pares, dissertações, teses e mais recentemente os dados de investigação e livros científicos (Swan, 2012), além de manuais, *software*, modelos tridimensionais, plantas, mapas, entre outros.

As barreiras de preço e de direitos de autor causam danos aos autores por limitarem a audiência e impacto do seu trabalho ao condicionarem o que pode ser lido, usado e citado (Suber, 2012, tradução livre). O movimento de AA veio ameaçar o modelo dominante de publicação científica, baseado até então em subscrições de revistas científicas de elevado custo, mudando o paradigma em termos de pagamento, isentando o leitor e cobrando ao autor. O AA não é acesso sem custos (OECD, 2015a). Em 2008, Stevan Harnad e Peter Suber propuseram termos para descrever os dois tipos de livre acesso à informação científica que se impuseram como fundamentais e reconhecidos. O termo Grátis refere-se à remoção das barreiras de preço, enquanto o termo Livre refere-se à retirada das barreiras de preço e de pelo menos algumas barreiras de permissão (Suber, 2008, tradução livre).

A viabilidade económica das infraestruturas de suporte ao AA pode constituir outro obstáculo. Contudo, as revistas de AA podem optar por vários modelos de financiamento que lhes garanta a sustentabilidade económica, como “subsídios, publicidade, taxas de pós-processamento do artigo, impressões, outros serviços de publicação, taxas de associado, ou combinações destas ações” (Universidade do Minho, 2017). Já os repositórios de AA são mais limitados em termos de sustentabilidade, no entanto, têm menos custos de manutenção e estão já em grande parte construídos (OECD, 2015a). O retorno do investimento que cada instituição faz num repositório tem sempre garantia de retorno através do conhecimento que aloja, atraindo desta forma alunos, investigadores e outros potenciais interessados.

A complexidade do AA reside na transição entre o modelo tradicional e o novo modelo de abertura, não pelos obstáculos legais, técnicos ou financeiros, mas sobretudo pelo fator cultural (Suber, 2012, tradução livre). O mesmo autor defende que o AA não é acesso universal, pois mesmo removendo as barreiras de preços e de permissões, continuam a existir outras barreiras como filtros e censura, barreiras de linguagem, de acesso a incapacitados e de conectividade.

A escolha do local de publicação por parte do investigador é muitas vezes influenciada pelo prestígio da revista, na procura da melhor visibilidade para a sua investigação. O prestígio das revistas é conseguido através do Fator de Impacto (FI), uma medida de avaliação criada por Eugene Garfield e Irving H. Sher: “In addition to helping libraries decide which journals to purchase, JIFs are also used by authors to decide where to submit their articles. As a general rule, the journals with high impact factors include the most prestigious, Some would equate prestige with high impact” (Garfield, 2006:92). O FI do ano de uma revista científica é calculado através da divisão

entre o número de citações total que os artigos dessa revista receberam nos dois anos anteriores e o número total de artigos publicados no mesmo período (Garfield, 2006). Suber afirmava em 2009 que o FI como medida principal da qualidade da investigação impede o crescimento das revistas de AA, pois, segundo o autor, pode afastar os investigadores da publicação em revistas de AA. Contudo, em 2017 existem já revistas de AA com o mesmo nível de prestígio das revistas tradicionais, das puras, como é o caso daquelas publicadas pela Public Library of Science (PLOS) e BioMed Central ou Hindawi às híbridas, de outros editores científicos.

Como incentivos para abertura dos seus trabalhos, além do maior reconhecimento e impacto medido pela contagem de citações, os investigadores apenas recebem recompensas menos tangíveis como a promoção na carreira (Suber, 2012). Para este autor, o sucesso do AA não depende do altruísmo do investigador, entendendo que o ganho de audiência e de impacto através da publicação da investigação em AA é do interesse pessoal do investigador.

Sendo o AA dirigido à literatura científica revista por pares seria de esperar que o controlo de qualidade não constituísse um problema. Contudo, surgiram no mundo editorial as revistas predatórias, um termo cunhado por Jeffrey Beall para definir as revistas que aproveitam o modelo dourado do AA em proveito próprio. “These predatory publishers are dishonest and lack transparency”. Constroem sítios web semelhantes aos dos “legitimate online publishers, and publish journals of questionable and downright low quality” (Beall, 2012). Beall assevera que “now there is a journal willing to accept almost every article, as long as the author is willing to pay the fee” (Beall, 2012). Este tipo de revistas cobra pela publicação de cada artigo sem, no entanto, oferecer os mesmos serviços editoriais como a revisão por pares, um elemento fundamental para garantir a integridade e qualidade científica.

O AA acarreta vários benefícios para a sociedade e comunidade científica como o aumento da velocidade, eficiência e eficácia da investigação, facilita a investigação interdisciplinar, amplia a visibilidade, uso e impacto da investigação e permite que as comunidades de profissionais, empreendedores e o público em geral, retirem valor da investigação (Swan, 2012, tradução livre). O projeto Study of Open Access Publishing (SOAP) concluiu que cerca de 90% dos investigadores que responderam ao inquérito consideram que o AA é benéfico para a investigação científica, com impacto direto positivo no fluxo da comunicação académica (Dallmeier-Tiessen et al., 2011).

De acordo com Harnad e Brody (2004), a melhor forma de calcular o impacto do AA é “compare the citation counts of individual OA and non-OA articles appearing in the same (non-OA) journals. Such ongoing comparisons are revealing dramatic citation advantages for OA”. Os mesmos autores enfatizam que o acesso não é uma condição suficiente para a citação, mas é uma condição necessária. Com as publicações em AA a crescer exponencialmente, entre as revistas de AA e os repositórios “verdes”, é natural

que o uso e impacto potencial aumentem, com variações entre campos de conhecimento decorrentes da sua cultura de partilha.

Após a análise de 36 estudos sobre o impacto do AA nas citações de artigos, Swan (2010) concluiu que 27 desses estudos apresentavam uma clara subida do número médio de citações. Os artigos científicos disponibilizados em AA são mais citados que aqueles que permanecem fechados, estando este facto relacionado com o aumento de audiência e visibilidade promovida pelo AA (Suber, 2012; Caruso, Nicol, & Archambault, 2013). Swan (2012) defende que o AA está a nivelar a comunicação científica e o acesso a ela. A mesma autora realça que a investigação científica tem impacto também fora do meio académico, ao nível empresarial, saúde ou educação. A OECD (2015a) acrescenta que os artigos são mais citados porque possuem livre acesso e uso e não propriamente pela sua qualidade. Um estudo da Science-Metrix concluiu que o AA pode aumentar o número de artigos citados e diminuir a citação de artigos que os investigadores não leram. Porém, a mesma análise revelou que apenas 8 dos 28 países auscultados conseguiram publicar 50% ou mais dos artigos produzidos em AA (Caruso, Nicol, & Archambault, 2013).

A distinção entre Grátis e Livre não possui o mesmo significado que a distinção entre via verde e via dourada. Enquanto a primeira distinção se refere aos direitos cedidos aos utilizadores (Quão aberto está?), a segunda distinção refere-se ao modo de publicação (Como é entregue?) (Suber, 2012). Os direitos de autor podem constituir uma barreira de peso ao AA. A abertura requerida pelo estatuto de AA Livre deve ser assegurada pela concessão de licenças que informem os utilizadores sobre o que podem e não podem fazer com determinado produto de investigação científica. Existem licenças abertas, como as Creative Commons (CC), que permitem a criação de modelos de licenciamento com diferentes níveis de permissões, podendo desta forma o detentor dos direitos de autor autorizar o nível de permissão que lhe seja mais conveniente.

Das 9431 revistas científicas de AA listadas no Directory of Open Access Journals (DOAJ) em Maio de 2017, 8447 dessas revistas utilizavam licenças CC, enquanto 278 revistas possuíam licenças próprias das editoras. Os cerca de 90% do número total de revistas de AA a utilizar licenças CC representa um crescimento exponencial desde 2012, quando Suber afirmou que apenas 20% das revistas em AA usavam algum tipo de licença CC. Nesta mesma data encontram-se listadas no DOAJ 78 revistas de AA de editoras portuguesas, sendo que 91% dessas publicações usa licenças CC para permitir o AA Livre (Directory of Open Access Journals [DOAJ], 2017).

A abertura dos resultados de investigação é suportada por políticas que podem encorajar ou requerer de forma obrigatória o AA. Por norma, o encorajamento é feito na via dourada, enquanto na via verde se definem mandatos. Segundo Swan (2012, tradução livre), na formulação de uma política de AA devem ser considerados os

seguintes fatores: Modo de publicação (via verde ou dourada); Local de depósito; Tipos de conteúdos abrangidos; Períodos de embargo; Permissões; Níveis de aplicação; Sensibilização de apoio às políticas; Sanções por incumprimento e Isenções.

De acordo com Suber (2012), existe similitude na abordagem às políticas de AA entre universidades e agências de financiamento, que tanto podem encorajar a via dourada como requerer a via verde. A Wellcome Trust do Reino Unido e a National Institutes of Health (EUA) são dois casos de sucesso de mandatos verdes, requerendo que os investigadores depositem o seu trabalho num repositório de AA e que retenham o direito de autorizar esse AA, o que permite que as instituições disponibilizem de forma aberta todos os resultados de investigação por si financiados (National Institutes of Health [NIH], 2008; Wellcome Trust, 2013).

O mais recente programa de financiamento de ciência e inovação da União Europeia, o Horizonte 2020, detalha no artigo 29.2 do *Model Grant Agreement* os requisitos legais na abertura das publicações científicas financiadas por este programa, onde cada beneficiário deve assegurar o AA a todas as publicações revistas por pares, utilizando para isso as duas vias disponíveis (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016a, tradução livre).

Existem também serviços com o intuito de facilitar a abertura das publicações científicas. O Registry of Open Access Repository Mandates and Policies (ROARMAP) lista as políticas e mandatos de AA ao nível global registados neste diretório. Em novembro de 2016, o ROARMAP conta com 794 mandatos instituídos por vários tipos de organizações. Portugal conta com 21 políticas de AA por via verde distribuídas na sua maioria por instituições de ensino superior. Os serviços SHERPA da Universidade de Nottingham, através das bases de dados RoMEO e JULIET, dispõem de valiosos indicadores quanto às políticas de AA das editoras científicas e das entidades financiadoras de investigação. Relativamente às políticas editoriais de AA, tendo em conta os direitos de autor e o auto arquivo, existem em Portugal 127 políticas segundo o SHERPA/RoMEO, sendo que mais de metade autoriza a publicação em AA do pós-print ou da versão final do editor. Já em termos de políticas de AA requeridas pela instituição financiadora, apenas a Fundação para a Ciência e Tecnologia (FCT) aparece listada no SHERPA/JULIET.

O AA envolve vários intervenientes distintos que cooperam na disponibilização dos resultados da investigação científica. Assim, como intervenientes podemos encontrar os investigadores, cujo esforço permite a criação e difusão de novo conhecimento com valor científico, a sociedade em geral que pode usufruir e construir em cima do novo conhecimento, as universidades e institutos de investigação que disponibilizam condições adequadas para o decurso da produção científica incluindo a construção de infraestruturas de AA, as entidades financiadoras de investigação científica, as editoras que criam revistas de AA para permitir a difusão acesso ao conhecimento e ainda

diversas organizações que dedicam parte das suas atividades à promoção de políticas e práticas de AA. Entre as organizações mais relevantes podemos encontrar a UNESCO, Wellcome Trust, a Joint Information Systems Committee (JISC), a Scholarly Publishing and Academic Resources Coalition (SPARC), a Open Knowledge International (OKI) ou a NIH, só para citar algumas das organizações mais reconhecidas pelo seu esforço em direção ao AA.

Em Portugal, os grandes intervenientes do AA são a FCT enquanto entidade financiadora, a Fundação para a Computação Científica Nacional (FCCN) no planeamento e gestão da Rede Ciência, Tecnologia e Sociedade (RCTS), a Universidade do Minho, pioneira no AA no país e o Repositório Científico de Acesso Aberto de Portugal, vulgo RCAAP, que é um agregador da informação científica produzida.

Segundo Suber (2012, tradução livre), o AA é um movimento amplo que abrange todos os campos de conhecimento, todos os países independentemente da sua realidade económica, inclui a investigação pública e privada sem limitação temporal e não se limita aos conteúdos textuais, pois frequentemente estende-se aos dados, conteúdos multimédia e código executável. A UNESCO (2015) defende que o AA não é um movimento isolado, embora tenha características únicas.

1.1.1 Via Dourada

A via dourada do AA é a publicação de artigos científicos em revistas de AA, que pode envolver o pagamento de custos de processamento - Article Processing Charge (APC) - pelo autor, instituição ou entidade financiadora, ficando o artigo disponível e sem custos para os leitores. A publicação em AA por via dourada é compatível com a qualidade científica das editoras tradicionais, pois as revistas de AA possuem o serviço de revisão por pares dos artigos, sendo que algumas delas já atingiram um alto nível de prestígio. O prestígio das revistas/editoras científicas é um fator que pesa na escolha do local de publicação por parte dos investigadores. Contudo, o mundo editorial de AA conta já com editoras de grande qualidade, que implementaram sistemas rigorosos de revisão científica, como são o caso da BioMed Central ou PLoS (Swan, 2012). O Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3) é um exemplo de sucesso de publicação em AA dourado. Sob orientação do European Organization for Nuclear Research (CERN), apostou no redirecionamento do dinheiro aplicado nas subscrições de revistas científicas para efetuar o pagamento dos APC's das revistas de AA da área da física das partículas (Sponsoring Consortium for Open Access Publishing in Particle Physics [SCOAP3], 2014). Algumas revistas científicas utilizam um modelo híbrido que lhes permite dobrar os lucros ao cobrar subscrições para acesso aos artigos, permitindo, contudo, que estes fiquem em AA se for pago o APC pelo autor (Swan, 2012).

O Reino Unido, um dos líderes mundiais em investigação, concluiu que a implementação de uma licença nacional pode potenciar o AA, ao permitir a negociação de preços ao mesmo tempo que orienta os investigadores para a publicação em AA dourado (Price & Chaytor, 2015).

1.1.2 Via Verde

De acordo com Harnad *et al.* (2004), a via verde de AA é um modo de publicação em revistas científicas de subscrição e o respetivo auto arquivo do trabalho publicado num repositório de AA. O auto arquivo é o ato de submeter uma cópia de um documento científico num repositório de AA de forma a assegurar a abertura e acesso ao seu conteúdo (UNESCO, 2015). Existem três tipos de documentos usados na via verde: a versão pré-publicada (manuscrito não sujeito a revisão), a versão pós-publicada (artigo aceite para publicação) e a versão final publicada (Laakso, 2014; Björk, Lakso, Welling, Paetau, 2014). Laakso (2014) sublinha a diferença de quantidade de artigos publicados em AA e indexados pela SCOPUS entre as várias disciplinas científicas, com a física e as ciências da saúde a serem as áreas do conhecimento mais abertas por larga margem. A falta de cultura de utilização do pré-print, a ausência de revistas científicas de alta qualidade ou a falta de financiamento para pagar os APC's são algumas das razões encontradas por Björk, *et al.* (2014) para justificar a diferença.

Os repositórios de AA são coleções ou bases de dados bibliográficas de artigos científicos publicados na web, mas incluem, outros conteúdos como teses, dissertações, *datasets* ou material didático, oferecendo uma ligação persistente a assegurando a preservação a longo-prazo (Suber, 2012). Os repositórios não efetuam o serviço de revisão por pares, embora alojem artigos publicados em revistas de AA. Outro aspeto importante da via verde de AA é o facto dos repositórios de AA não poderem gerar permissões de utilização, por geralmente não possuírem esses direitos (Suber, 2012; Björk, et al., 2014).

Segundo Houghton & Swan (2013), a via verde é o meio mais sustentável em termos de custos para todas as partes envolvidas. Porém, de acordo com Swan (2012), só quando o AA por via verde é mandatado é que se atingem resultados satisfatórios. Apesar disso, os repositórios podem estimular a sua utilização se fornecerem serviços como a contagem de citações ou estatísticas de uso. Bjork *et al.* (2014, tradução livre) afirmam que o rácio de depósito de artigos revistos pelos pares em repositórios ronda os 12%, justificando este baixo número com os fracos sistemas de incentivo académico, com a falta de repositórios adequados, com a falta de conhecimento e com o fato de ser uma tarefa que consome tempo.

Segundo o Directory of Open Access Repositories (OpenDOAR), existem vários tipos de repositórios de AA: os disciplinares que reúnem toda a informação científica de um

determinado campo do conhecimento, os institucionais, que agregam a informação produzida por uma determinada instituição, os governamentais que alojam informação proveniente de órgãos do Governo e os repositórios agregadores. Em Novembro de 2016 encontram-se listados no OpenDOAR 49 repositórios de AA localizados em Portugal, sendo que cerca de $\frac{3}{4}$ da totalidade dos repositórios pertence a instituições do ensino superior. O Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) permite a interoperabilidade entre diferentes plataformas, possibilitando aos repositórios de AA formar um sistema central virtual, potenciando a descoberta e acessibilidade da informação científica (Suber, 2012; Swan, 2012; Björk, et al., 2014). A Cybermetrics Lab lançou em 2017 o Ranking de Repositórios Web, um grupo de investigação do Conselho Superior de Investigação Científica de Espanha, com o objetivo de apoiar o AA às publicações científicas.

Portugal possui 4 repositórios no grupo dos 223 “Top Portals” do ranking, com o Scielo Portugal a ocupar a 76ª posição, o Repositório Comum na 87ª posição, a Plataforma de Revistas de AA da Universidade de Aveiro na 110ª posição e o Depósito de Dissertações e Teses Digitais na 124ª posição. No “Top Institutionals”, Portugal é representado pelo Universidade do Minho Repositorium na 60ª posição, seguido pelo Repositório Universidade de Lisboa na 156ª posição e Repositório Aberto Universidade do Porto na 188ª posição, entre os melhores duzentos classificados. Em termos de ranking mundial, o Repositorium da Universidade do Minho é o repositório melhor colocado, ocupando a 68ª posição (Cybermetrics Lab, 2017).

A relação entre a via verde e a via dourada pode ser vista como “complementary and synergistic”, pois a filosofia de criação e implementação é a mesma (Suber, 2012:58). Para este autor, a via verde não necessita de investimento económico tão elevado para alavancar o AA às publicações científicas nem infringe a liberdade académica, podendo ainda cobrir todos os resultados de investigação de uma instituição. Este modo de disseminação é compatível com a publicação nas tradicionais revistas por subscrição, podendo ainda alojar outros conteúdos que não apenas artigos. Já a via dourada possui a vantagem de ser imediata, sem períodos de embargo, podendo ainda ser classificada como AA Livre, se forem dadas as respetivas permissões de uso. Acrescenta-se o facto de a via dourada possuir o serviço de revisão por pares e fornecer sempre o acesso à versão final do artigo. Suber (2012, tradução livre) afirma que apenas a conjugação das vantagens de cada modelo de publicação garante a complementaridade de publicação em AA, uma vez que por si só, tanto a via verde como a via dourada não são suficientes para assegurar o pleno AA. O autor defende que enquanto a via verde atua melhor no registo e preservação da investigação científica, a via dourada assegura uma melhor distribuição e certificação de qualidade. Na impossibilidade de publicar por uma das vias, o investigador pode optar pela outra via, assegurando sempre o AA ao seu trabalho.

1.2 A Ciência Aberta

A investigação científica encontra-se numa transição cultural que abrange os modelos de publicação e reutilização na procura de uma maior abertura, com o cariz digital a impulsionar o acesso aos resultados científicos. Machado (2015) salienta a necessidade de adotar um conceito mais amplo e que possa alojar na sua definição toda a panóplia de recursos científicos produzidos durante o processo de investigação. A Ciência Aberta (CA) refere-se comumente ao esforço para tornar mais acessíveis os resultados da investigação científica financiada pelo setor público, em formato digital, para a comunidade científica, setor empresarial ou sociedade em geral (OECD, 2015a).

O conceito de CA é considerado um termo guarda-chuva que assenta em vários pilares quer para remoção das barreiras de partilha de qualquer resultado, método ou ferramenta em qualquer fase do processo de investigação (Bueno de la Fuente, 2016, tradução livre), quer como sua condição para a sua concretização (infraestrutura tecnológica), quer como desambiguação de autores, instituições e produção científica (identificadores únicos). A figura seguinte expressa algumas das vertentes que se enquadram no conceito amplo de CA e que respeita a manifestações muito distintas, não apenas da academia, como já foi referido, mas de todas as esferas da sociedade.

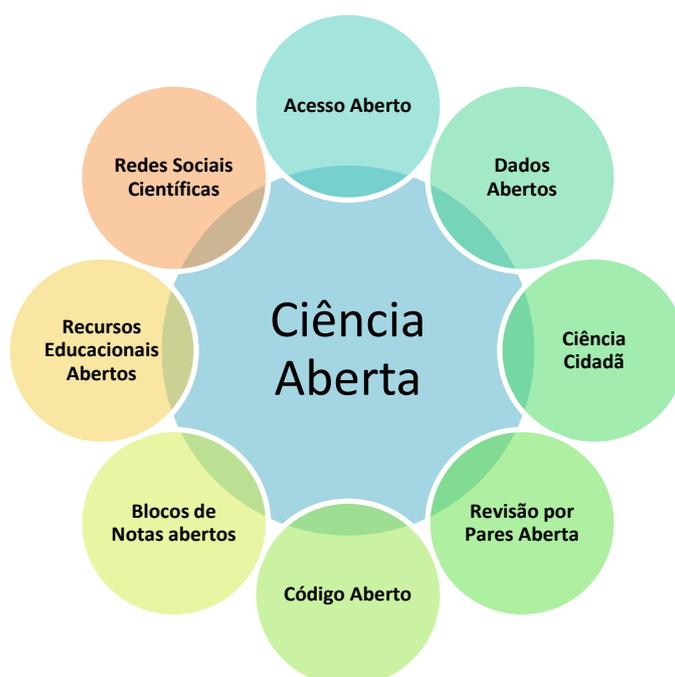


Fig. 1 Os pilares da Ciência Aberta. Fonte: adaptado de <https://www.fosteropenscience.eu/content/what-open-science-introduction>

A CA é um meio e não um fim. Garantir o acesso aos resultados de investigação científica pode tornar o sistema de investigação mais eficiente ao reduzir custos na coleta, transferência e reutilização de dados, ao aumentar a transparência e qualidade no processo de investigação, acelerar a transferência do conhecimento, aumentar a

repercussão do conhecimento na economia e promover a ligação entre cidadãos e a ciência e a investigação (OECD, 2015a, tradução livre). A colaboração dos cidadãos no processo de investigação é denominada ciência cidadã, sendo o movimento de CA um importante facilitador da participação pública na ciência (Leonelli, Spichtinger, Prainsack, 2015). O crescimento da ciência cidadã aumentou o número de intervenientes na cocriação de conhecimento, contribuindo para a aproximação da fronteira entre o cientista profissional e o amador (Royal Society, 2012; International Science, 2015). Segundo a UNESCO (2012), citando dados da PubMed Central, 40% do acesso a resultados de investigação é feito por cidadãos individuais, superando a utilização feita por universidades, empresas e governo.

Segundo a Open Science and Research Initiative (2014), a promoção da CA é feita em 4 níveis: definição de políticas gerais, alteração na cultura de trabalho, alteração nos métodos de trabalho e desenvolvimento de serviços e infraestruturas. Fecher & Friesike (2013) estruturaram as mudanças na CA em cinco escolas de pensamento:

Infraestrutura	• dirigida à arquitetura tecnológica, com novas plataformas, ferramentas e serviços para os cientistas
Pública	• dirigida à acessibilidade do conhecimento pelo público
Avaliação	• dirigida ao desenvolvimento de uma nova avaliação alternativa de impacto
Democrática	• dirigida ao acesso gratuito ao conhecimento pelo público
Pragmática	• melhoria da eficiência através da investigação colaborativa

Fig. 2. As 5 escolas de pensamento das mudanças operadas pela Ciência Aberta. Fonte: Fecher & Friesike (2013).

A abertura aos resultados da investigação é, segundo a UNESCO (2015), muito requerida no processo de capacitação dos cidadãos, sendo que esta noção de abertura acarreta mudanças na sociedade ao fornecer soluções sustentáveis para o chamado fosso digital⁶. Para a CA prosperar, a comunidade académica deve ter acesso a publicações, dados, métodos, conhecimentos, serviços de apoio e colaboração entre os diversos intervenientes (Open Science and Research Initiative, 2014; Leonelli, Spichtinger, Prainsack, 2015; OECD, 2015a; Resolução do Conselho de Ministros n.º 21/2016, 2016). O acesso à literatura académica tem por objetivo transpor a clivagem de conhecimento existente na sociedade, fornecendo o Acesso Aberto (AA), através de sistemas em linha, aos resultados da investigação financiada com fundos públicos (UNESCO, 2015).

⁶ Segundo o glossário estatístico da OCDE, o termo fosso digital refere-se à desigualdade no acesso à informação, tecnologias de comunicação e uso da Internet entre indivíduos, negócios e áreas geográficas com diferentes níveis socioeconómicos. Disponível na WWW: <https://stats.oecd.org/glossary/detail.asp?ID=4719>

O projeto da EU Facilitate Open Science Training for European Research (FOSTER) assume a divisão do conceito de Ciência Aberta em quatro diferentes categorias: Acesso Aberto, Dados Abertos, Código Aberto e Investigação Replicável Aberta (Pontika, Knoth, Cancellieri, & Pearce, 2015).

A plena implementação da CA está dependente da resolução de alguns desafios. A Royal Society (2012) afirma que os interesses comerciais e económicos incluem a propriedade dos dados e os direitos de propriedade intelectual associados, as parcerias público-privadas e a abertura de informação comercial de interesse público, a que se junta a privacidade e a segurança e que constituem limites à implementação da CA. Por outro lado, os incentivos e as recompensas concedidos à comunidade científica continuam a ser inexistentes, particularmente porque existe: (i) falta de infraestruturas para apoiar a abertura, (ii) receio que os dados em bruto e métodos sejam mal interpretados ou mal usados, (iii) incerteza sobre a propriedade dos dados e (iv) falta de experiência na promoção da CA (Open Science and Research Initiative, 2014, tradução livre). Metaforicamente, Leonelli, Spichtinger, Prainsack (2015:13) defendem que “in addition to the stick of compulsory mandates, carrots are therefore also needed”. Os autores sintetizam as “carrots”, ou seja, as recompensas, em três eixos: reconhecimento das práticas de partilha nas estruturas de incentivos, criação de incentivos mais significativos para envolver os investigadores na CA e o reconhecimento do papel das métricas alternativas na avaliação dos investigadores.

A CA proporciona vários benefícios a todos os intervenientes e partes interessadas do processo de investigação, como o aumento da eficácia da investigação, o aumento do conhecimento do modelo científico, melhoria do foco e da qualidade dos resultados de investigação, geração de novas ideias de investigação, aumento do compromisso com a ciência e melhoria da literacia científica, maior abertura e impacto social e económico, aumento da qualidade, visibilidade e impacto da investigação (Open Science and Research Initiative, 2014; Leonelli, Spichtinger, Prainsack, 2015; Forsström & Haataja, 2016). Estes benefícios abrangem os investigadores e as suas equipas, as organizações, os decisores e financiadores e também o público em geral.

A CA clama por uma mudança no comportamento dos investigadores de forma a permitir a abertura de conteúdos e a adoção de práticas e estratégias abertas em todo o ciclo de investigação (Pontika et al., 2015). De acordo com Machado (2015), a abertura do conhecimento na sociedade da informação pode ser uma realidade se essa for a prioridade de qualquer política que tenha por objetivo promover a melhoria das condições de vida dos cidadãos e a redução das assimetrias globais. Quando falamos de CA, aquilo que se vê é uma realidade já espelhada globalmente, inclusivamente em países como a China (Price & Chaytor, 2015), onde a abertura política não o faria supor.

A Comissão Europeia desenvolveu o Open Science Monitor como forma de avaliar os desenvolvimentos na CA e as tendências particulares “in open science activities over time and comparatively between countries and scientific disciplines” (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2017). O Open Science Monitor tem como fonte diversas organizações empenhadas na avaliação da CA, do AA e dos Dados Abertos⁷, garantindo assim a profundidade e qualidade dos dados. O modelo usado pelo Open Science Monitor divide diversos indicadores de avaliação em três categorias: AA às publicações, AA aos dados de investigação e formas abertas de comunicação académica, como demonstrado na figura⁸ seguinte:

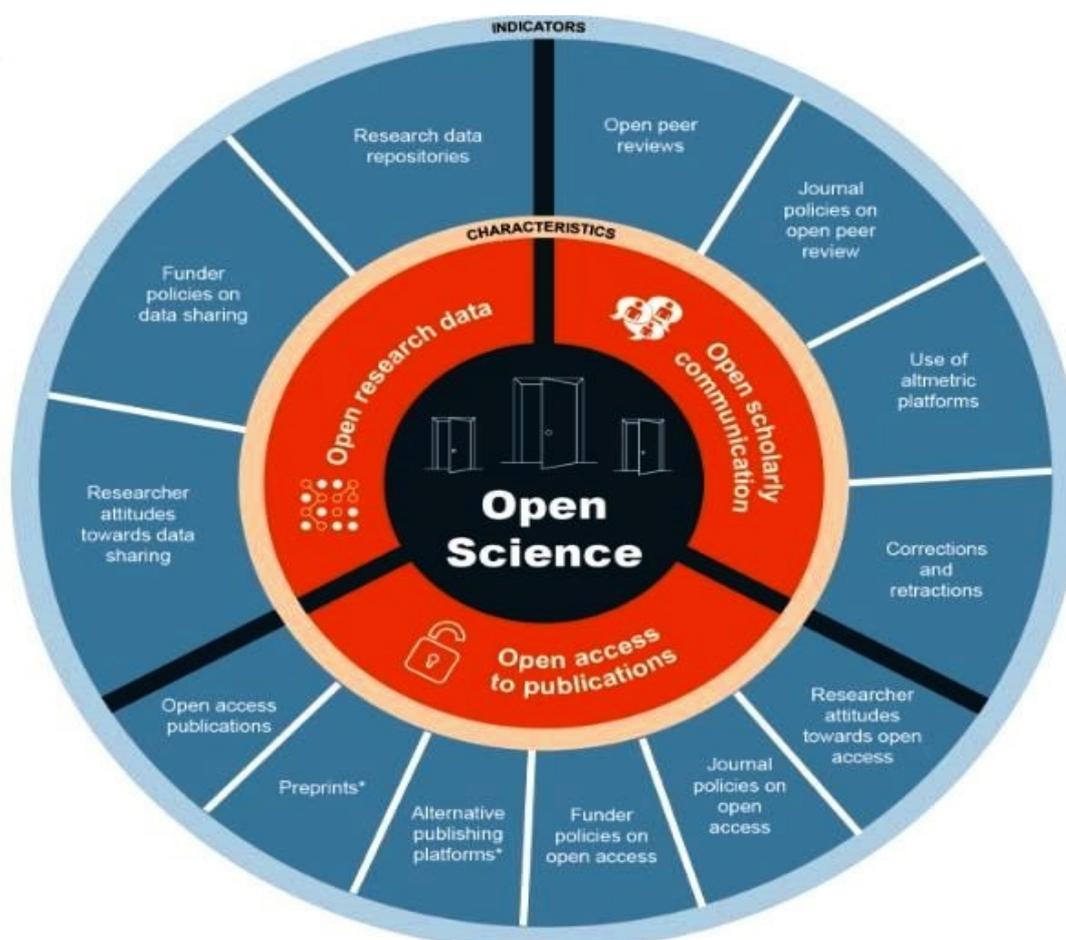


Fig. 3. Características e indicadores do Open Science Monitor. Fonte: Comissão Europeia: Diretoria Geral para a Investigação e Inovação (2017). Disponível na WWW: <http://ec.europa.eu/research/openscience/index.cfm?pg=home§ion=monitor>

A implementação da CA implica alterações ao nível da administração da investigação para melhorar a ciência, tornando-a credível, replicável e íntegra, como explica a Comissão Europeia (2016a). Significa também fazer alterações técnicas e desenvolver normas para tornar a ciência mais eficiente ao evitar a duplicação de

⁷ Entre as várias organizações encontram-se a 101 innovations, Clarivate Analytics, Altmetric, arXiv, F1000Research, figshare, OpenAIRE, Wiley, entre outras.

⁸ Os indicadores assinalados com um asterisco (*) também são considerados como forma de comunicação académica aberta.

recursos e ao otimizar a reutilização de dados. Por fim, tornar a ciência mais aberta ao melhorar a acessibilidade a dados e ao assegurar as condições legais apropriadas para o desenvolvimento de *Text e Data Mining*, potenciando assim alterações económicas e legais ao nível do copyright.

Os investigadores recorrem cada vez mais às plataformas sociais para partilhar e validar os projetos de investigação, enquanto a criação de plataformas abertas permite a partilha de dados para resolução de problemas públicos, como é o caso do The Humanitarian Data Exchange⁹ (Open Science and Research Initiative, 2014). A ScienceOpen é outro exemplo de uma plataforma gratuita de pesquisa de informação científica que oferece contexto à investigação através de comentários, recomendações ou revisão por pares após a publicação, permitindo assim que os investigadores encontrem a informação mais relevante para as suas pesquisas (ScienceOpen, 2017). Nos últimos anos têm surgido outras formas de publicação em AA, normalmente através de páginas pessoais ou em sítios de comunidades científicas como o academia.edu ou o Research Gate, que em 2017 foram classificados como líderes do ranking “Top Portals” do Ranking de Repositórios Web pela Cybermetrics Lab. Outros resultados interessantes foram obtidos por Kramer e Bosman (2016) após a análise dos dados dos mais de 20 mil participantes num inquérito a nível global. A CA e o AA são apoiados por 80.9% e 85.9% dos inquiridos respetivamente. Em termos de publicação de artigos, o estudo indica o Research Gate e os repositórios institucionais como os preferidos. Relativamente à publicação de dados, a preferência dos inquiridos recaiu no GitHub seguindo-se o Figshare, DropBox, BitBucket, Dataverse, Dryad e o Zenodo. Porém, dos 20 mil inquiridos apenas cerca de 30% respondeu a esta questão. Em relação a Portugal e no caso do apoio à CA e ao AA, os resultados são extremamente positivos, contudo, a amostra obtida é demasiado pequena para se poder aferir qualquer conclusão definitiva. E contrastam de forma gritante com as conclusões do relatório sobre a elaboração de uma política nacional de CA, que sublinha a dificuldade de generalização das práticas de AA, registando por isso limitações nas “taxas de sucesso e cumprimentos das políticas institucionais” (Comissão Executiva Grupo de Trabalho para a Política Nacional de Ciência Aberta [Comissão Executiva GT-PNCA], 2016:9).

A intenção da EU é criar uma European Open Science Cloud (EOSC), ou seja, um ambiente onde os investigadores possam ter acesso a ferramentas e dados para fins de investigação, inovação e educação. Esta plataforma é considerada uma parte fundamental da transição europeia para a CA e para tornar a ciência orientada por dados (Comissão Europeia, 2016a, tradução livre).

Podem observar-se algumas tendências chave na CA e a sua relevância na EOSC, como o novo modelo de comunicação académica: recompensas modernas e

⁹ Disponível na WWW: <https://data.humdata.org/>

reconhecimento, aumento da dependência de peritos de dados, colaboração interdisciplinar, fomento da transição de ciência para inovação, desenvolvimento de um complexo sistema de infraestruturas, da integridade da investigação e do processamento do conhecimento por máquinas (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016b, tradução livre).

O FOSTER desenhou uma interessante taxonomia a partir do termo CA, que permite demonstrar a grandeza e os diferentes níveis de impacto do conceito (Pontika et al., 2015).

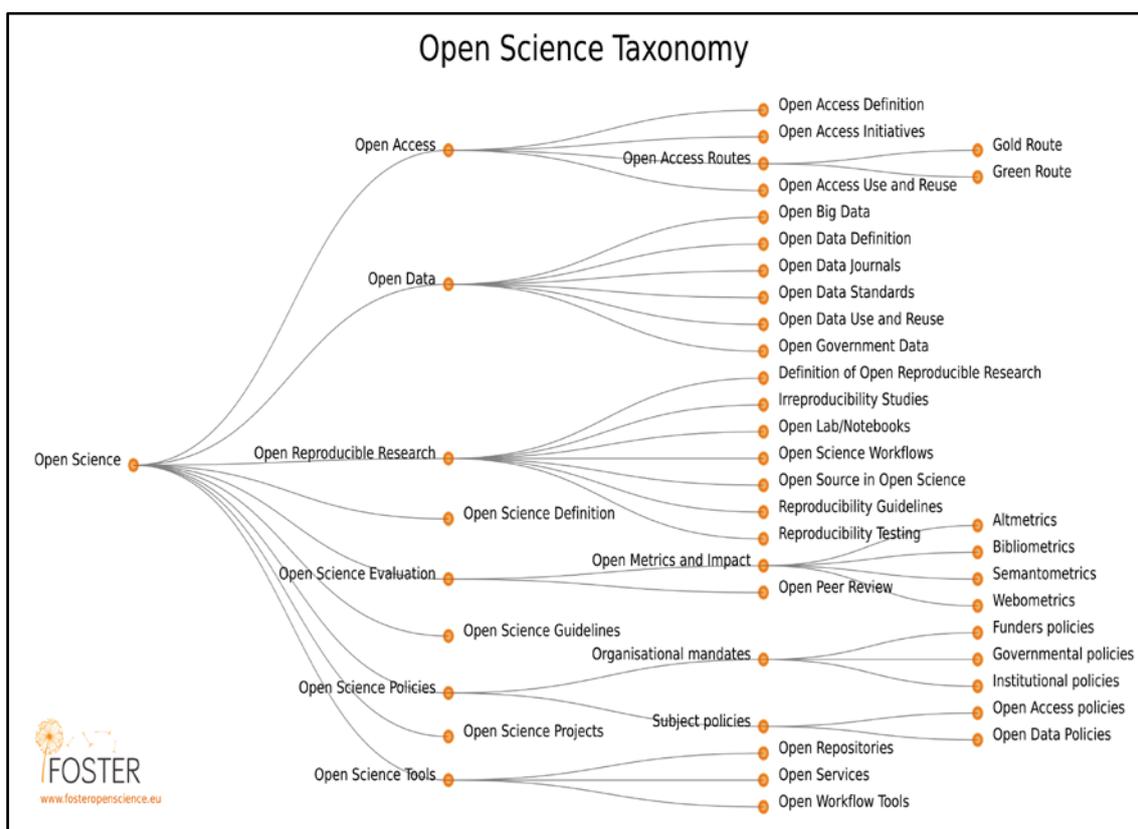


Fig. 4. Taxonomia da Ciência Aberta. Fonte: Knoth, Petr; Pontika, Nancy (2015).

Em termos de políticas, existem três tendências na definição das políticas de CA, como refere a OECD (2015a, tradução livre):

- As regras obrigatórias impostas nos contratos de financiamento ou em políticas nacionais;
- Os mecanismos de incentivo como o apoio financeiro para publicação de investigação, a citação e a progressão na carreira;
- Os viabilizadores, ou seja, a infraestrutura que permite o decurso da abertura científica.

A OECD (2015a) acrescenta que as políticas de AA são mais maduras que as políticas de Ciência Aberta existentes e sublinha ainda que as políticas de CA devem ter princípios-base mas que têm de ser adaptadas às realidades locais.

A CA tem estado em foco na agenda política dos governos e organizações mundiais. Ao nível europeu, a Agenda Digital da Comissão Europeia define cinco potenciais linhas de ação políticas (Comissão Europeia, 2016a:45, tradução livre):

- Fomento e criação de incentivos para a CA, através do fomento do conceito de abertura em programas de educação e promoção das melhores práticas de ciência cidadã, garantindo a qualidade, impacto e integridade da investigação;
- Remoção de barreiras através da revisão de carreiras dos investigadores bem como dos incentivos e recompensas pelos esforços de CA;
- Generalização da promoção de políticas de AA (artigos e dados de investigação);
- Desenvolvimento de infraestruturas de investigação para a CA de forma a melhorar o alojamento de dados, acesso e governação com a criação de uma nuvem europeia de CA (European Open Science Cloud);
- Incorporar a Ciência Aberta na sociedade como condutor socioeconómico.

De acordo com a Open Science and Research Initiative (2014), a CA tem vários intervenientes distintos: os investigadores e as equipas de investigação, organizações/entidades de investigação, financiadores, decisores e público em geral. A OECD (2015a) acrescenta as bibliotecas, editoras científicas privadas, empresas e entidades supranacionais, como a própria OECD. Neste contexto, a EU tem-se destacado como promotora da CA. Nesse sentido criou o já mencionado FOSTER, um projeto cujo objetivo é produzir um programa de treino que irá ajudar investigadores, estudantes, bibliotecários e outras partes interessadas a incorporar a CA nas suas metodologias de investigação. O portal FOSTER¹⁰, é uma plataforma *e-learning* que agrega conteúdos de treino sobre CA, para aqueles que precisem de desenvolver estratégias e competências para a implementação da CA nos fluxos de trabalho diário (FOSTER, 2016). A União Europeia (EU) posiciona-se também como financiadora do AA como forma de acesso ao conhecimento. Nesse sentido, adotou o AA como princípio em 2012, tendo-o definido como obrigatório em 2013 no novo programa de financiamento da investigação e inovação europeu, denominado Horizonte 2020 (União Europeia, 2013). Financiou ainda um projeto de investigação exclusivamente dedicado ao alinhamento de estratégias e políticas de AA, o Pasteur4OA. Além disso, a EU criou o OpenAIRE, uma plataforma que permite agregar toda a literatura científica produzida em projetos financiados por fundos europeus. A segunda fase do OpenAIRE consiste em facilitar o AA ao fornecer ligações cruzadas entre publicações, dados e esquemas de financiamento e irá expandir a rede de repositórios para incluir os fornecedores de dados com o objetivo de interligar os dados científicos associados

¹⁰ Disponível na WWW: <https://www.fosteropenscience.eu/>

(OECD, 2015a). A terceira fase do projeto arrancou em Janeiro de 2017 com o OpenAIRE-Connect que visa fornecer pontes tecnológicas e sociais, para além de disponibilizar serviços de troca de literatura, dados ou métodos de investigação “with semantic links between them, across research communities and content providers in scientific communication” (OpenAIRE, 2017). O OpenAIRE-Connect irá também implementar o conceito de “Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure, delivering out-of-the-box, on demand deployable tools” (OpenAIRE, 2017). O projeto é uma parceria entre 11 instituições onde consta a Universidade do Minho como representante portuguesa.

A UNESCO está ativamente empenhada na promoção da CA e em parceria com a Nature lançou, em Novembro de 2014, a Biblioteca Mundial de Ciência¹¹ (WLoS), uma biblioteca científica gratuita e multilingue com o objetivo de fornecer acesso à investigação científica recente a estudantes e professores de todas as partes do globo. Podemos ainda relacionar a CA com o conceito de Ciência 2.0 que sugere a partilha e a colaboração como nova abordagem para o uso de informação, incluindo o uso de ferramentas como *wikis*, *blogs*, *vídeos*, dados em bruto e novas teorias (Machado, 2015; Leonelli, Spichtinger, Prainsack, 2015).

Em Portugal, só em fevereiro de 2016 se regista a intenção de estabelecer uma política nacional de Ciência Aberta com a publicação dos princípios orientadores para a “Ciência Aberta, Conhecimento para todos” pelo Ministério da Ciência, Tecnologia e Ensino Superior (MCTES). Em Maio, a Resolução do Conselho de Ministros nº 21/2016 vem definir os princípios da implementação da política de Ciência Aberta em Portugal. Esta política está estruturada em torno de quatro eixos: Acesso aberto e dados abertos; Infraestruturas e preservação digital; Avaliação científica; Responsabilidade social científica (Comissão Executiva GT-PNCA, 2016).

O MCTES e o Ministério da Cultura lançaram em 2017¹² o Diretório dos Repositórios Digitais, uma infraestrutura digital com o objetivo de identificar e constituir um ponto de acesso aos repositórios digitais existentes em Portugal, na área da Ciência e da Cultura, de forma a promover a utilização dos repositórios digitais, bem como a normalização técnica e práticas de preservação digital, tal como foi anunciado em 2016 (Portugal. Ministério da Ciência, Tecnologia e Ensino Superior, Ministério da Cultura, 2016).

1.3 *Big Data* ou dados massivos

O termo *Big Data* surge associado aos dados em formato digital que excedem a capacidade de processamento dos sistemas de bases de dados convencionais (Reitano,

¹¹ Disponível na WWW: <http://www.nature.com/wls/about>

¹² Disponível na WWW: <http://repositorios-conhecimento.pt>

2013, Comissão Europeia, 2014). Segundo Borgman (2015), o *Big Data* compreende dados estruturados e não estruturados, com tendência à homogeneidade e à padronização. Ubaldi (2013, tradução livre) reitera que o Big Data enquanto fenômeno emergiu quando a disponibilidade de *datasets* produzidos por várias fontes cresceu e quando os utilizadores de dados reconheceram o valor potencial da combinação e ligação de diferentes conjuntos de dados produzidos por diferentes intervenientes.

A revolução digital ocorrida nas últimas décadas teve repercussões profundas na coleta, acesso, armazenamento e transmissão de dados, estando atualmente estimado em 40% o volume de crescimento de dados por ano (Manyika, et al., 2011). Os dados digitais estão presentes em todos os setores económicos, o número de páginas web cresceu exponencialmente, surgiram as redes sociais, os *smartphones* e toda uma panóplia de tecnologias que coletam dados que podem ser benéficos para a sociedade, como as aplicações surgidas no âmbito da Internet of Things¹³ (IoT) que irão avolumar o número de dados já produzidos (Fan & Bifet, 2012):

We're drowning in data, bits are faster than atoms (Croll, 2012:2).

Large amounts of data are being produced permanently, and when these are analyzed and interlinked they have the potential to create new knowledge and intelligent solutions for economy and society (União Europeia, 2016).

Lynch (2009) salienta a emergência do 4º paradigma da investigação científica, termo cunhado por Jim Gray, definindo-se como uma ciência intensiva orientada a dados.

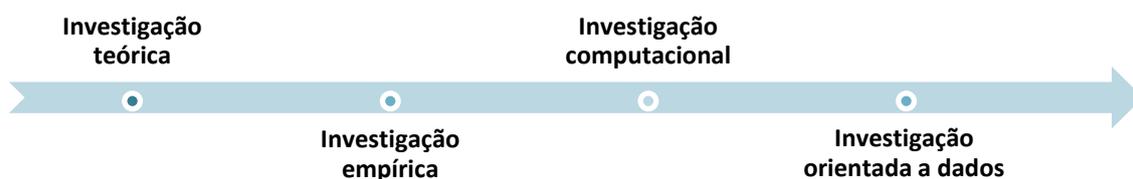


Fig. 5. Paradigmas da investigação científica. Fonte: traduzido de Lynch (2009)

Esta nova ciência assenta em três pilares: captura, curadoria e análise de dados (Bell, 2009, Royal Society, 2012). A curadoria de dados envolve variadas atividades, como “finding the right data structures to map into various stores” (Bell, 2009:XIII). Inclui o esquema de metadados e os respetivos metadados para a sua preservação e pesquisa. A análise de dados “covers a whole range of activities throughout the workflow pipeline, including the use of databases, analysis and modeling, and then

¹³ A Internet Of Things é um conceito onde edifícios, transportes, produtos e pessoas se tornam espaços de informação, ou seja, são imbuídos com sensores, *software* ou eletrónica que lhes permite coletar e partilhar dados. Disponível na WWW: <http://www.theinternetofthings.eu/what-is-the-internet-of-things>

data visualization” (Bell, 2009:XIV), podendo ser sinónimo de melhores resultados, processos e decisões (Comissão Europeia, 2014).

Dumbill (2012) afirma que para retirar valor destes dados é necessário encontrar formas alternativas de os processar. Borgman reitera que os dados têm muitos tipos de valor e que esse valor pode não ser visível até os dados serem coletados, curados ou perdidos. A autora sublinha que os dados não têm valor estando isolados, eles fazem parte de uma infraestrutura do conhecimento, “an ecology of people, practices, technologies, institutions, material objects and relationships” (Borgman, 2015:4). Este volume de dados produzidos pode, segundo o Big Data Europe (2016), trazer contribuições relevantes para o progresso técnico dos setores chave da sociedade e ajudar a reformular o mundo dos negócios. Ainda segundo a EU, “what is needed are innovative technologies, strategies and competencies for the beneficial use of Big Data to address societal needs”.

Segundo Manyika *et al.* (2011), os Dados Massivos podem gerar valor de várias formas: criar transparência governativa e empresarial; permitir a experimentação, expor variáveis e melhorar o rendimento; criar de produtos e serviços para populações específicas; apoiar ou substituir o processo de decisão humana com algoritmos automatizados; criar de modelos de negócio, produtos e serviços inovadores. Verifica-se que um número crescente de empresas já usa dados para gerar valor económico, sendo que também os governos têm a possibilidade de aumentar a sua eficiência e gerar poupança (Manyika, et al., 2011). A Comissão Europeia (2017a) destaca as aplicações para *smartphones* ou produtos inteligentes para empresas, acrescentando que os Dados Massivos podem fomentar o crescimento de emprego na Europa e melhorar a qualidade de vida dos cidadãos europeus.

Numa organização o valor do *Big Data* é contabilizado em duas categorias: uso analítico e desenvolvimento de novos produtos (Dumbill, 2012). Croll (2012:2) afirma a necessidade da sociedade em decidir, agir e experimentar com base nas aprendizagens feitas e utilizar esse conhecimento em ações futuras, defendendo que estamos a entrar numa “feedback economy” que irá tornar-se a norma em organizações públicas e no setor empresarial, dada a eficiência e a otimização de recursos geradas. Esta economia de retorno de resultados e respetivas medições de impacto tem por base a cadeia de abastecimento dos Dados Massivos, como representado na figura seguinte:



Fig. 6. Cadeia de abastecimento dos Dados Massivos. Fonte: traduzido de Croll (2012).

“In this networked world, people, objects and connections are producing data at unprecedented rates, both actively and passively” (Science International, 2015). Segundo Hey, Tansley e Tolle (2009), a Internet pode proporcionar mais do que o acesso aos dados, pode agregar todos os dados disponíveis à literatura que lhes faz referência, aumentando a velocidade da informação científica e a eficiência da investigação. Laney (2001) foi o primeiro a caracterizar a gestão de dados através do sistema de três V’s que ajuda a perceber a natureza dos dados e as plataformas disponíveis para os explorar. Marr (2014) afirma que atualmente o sistema passou a ser composto por dois V’s adicionais, perfazendo um total de cinco V’s:

- Volume: o volume e o tamanho dos dados continuam a aumentar, contudo a capacidade de processamento dos nossos sistemas continua baixa;
- Variedade: existência de diversos tipos de dados provenientes de diversas fontes;
- Velocidade: obtenção de informação em tempo real;
- Veracidade: refere-se à qualidade e correção dos dados;
- Valor: ganho de competitividade, tomada de decisões alavancada em informação atualizada.

A intensidade da revolução operada pelo fenómeno dos Dados Massivos “has produced changes in economic and social behaviour e has profound implications for science”, permitindo a identificação de padrões em fenómenos e demonstrando relações insuspeitas entre dados que até aqui eram desconhecidas (Science International, 2015). “Contrary to many high-volume big data settings, where standardised infrastructure are available, there exist other big data settings with less

mature processes, due to the lack of tools, resources and community exchange” (Pröl, Meixner & Rauber, 2016). Borgman (2015:8) recorre a um termo utilizado na Estatística para identificar distribuições de dados, a *long tail of data*¹⁴ ou cauda longa dos dados, que quando aplicada à investigação académica implica que “a small number of research teams work with large volumes of data, some teams work with very little data, and most fall somewhere in between”.



Fig. 7. A cauda longa dos Dados Massivos. Fonte: traduzido de Borgman (2015).

A cauda longa dos dados de investigação abrange grandes porções “of data that are highly heterogeneous, managed predominantly locally within each researcher’s environment, and frequently not properly transferred to and managed within well-curated repositories” (Pröl, Meixner & Rauber, 2016). Segundo o e-Infrastructure Reflection Group (e-IRG) (2016), citando Heidorn (2008), o termo *long tail of data* tem-se também aplicado para salientar a variedade na estrutura, tema, complexidade, formato, tamanho, localização e o uso na investigação dos dados de investigação.

Para capturar todo o potencial gerado pela produção massiva de dados devem ser tidas em conta várias questões (Manyika, et al., 2011, tradução livre):

- Formulação de políticas de dados que incluam temas sensíveis como a privacidade, segurança, credibilidade e propriedade intelectual;
- Desenvolvimento de novas tecnologias de armazenamento, computação e *software* analítico além de novas técnicas de análise;
- Mudança na cultura organizacional com a contratação de peritos que possam retirar informação e conhecimento dos dados e também o estabelecimento de incentivos para otimizar o uso do *Big Data*, induzindo melhorias numa tomada de decisão mais informada;
- Acesso aos dados para permitir a criação de novas oportunidades de negócio;
- Intensificação da estrutura competitiva de empresas e setor público para potenciar a captação de benefícios da massificação de dados.

¹⁴ O termo foi introduzido por Cris Anderson (2004) para comparar o mercado de bens feito ao nível de vendas físicas e *online*. Disponível na WWW: <https://www.wired.com/2004/10/tail/>

O fenômeno do *Big Data* está intrinsecamente ligado a outro fenômeno, o desenvolvimento de *software* aberto como é o caso do Apache Hadoop¹⁵, que permite o processamento de grandes quantidades de dados em paralelo através “large clusters of compute nodes” (Fan & Bifet, 2012:3).

A influência do *Big Data* na sociedade manifesta-se a diversos níveis, destacando-se o desenvolvimento de projetos na área da saúde, transportes, agricultura, ambiente, indústria, ganho de competências científicas, mercado de dados e dados abertos. Num artigo assinado na *Computer World*, Olavsrud (2013) identifica alguns dos usos do *Big Data* na atualidade: fins médicos, empresariais, tecnológicos, em eventos desportivos ou no desenvolvimento das *smart cities*. A União Europeia tem contribuído para o alargar da importância e da esfera de influência do fenômeno *Big Data*, financiando, entre outros, projetos como:

- DataPitch¹⁶ que tem por objetivo acelerar a chegada dos dados ao mercado;
- Open Data Incubator Europe¹⁷ (ODINE) que oferece incubação e tutoria às PMEs na nova geração de empreendedorismo digital;
- European Data Science Academy¹⁸ (EDSA) que avalia o estado do setor de peritos de dados e oferece ciclos de aprendizagem de habilidades específicas, de acordo com as necessidades detetadas;
- Big Data Europe que visa coletar informações sobre o desenvolvimento de infraestruturas necessárias à prática de uma ciência de dados intensiva para a resolução de problemas da sociedade, possibilitando que as empresas possam construir produtos e serviços inovadores “based on semantically interoperable, large-scale, multi-lingual data assets and knowledge, available under a variety of licenses and business models” (Big Data Europe, 2016).

1.4 Dados Abertos

É axiomático o fato de o conhecimento ser essencial ao processo de decisão, à inovação e ao bem-estar social. Os dados sempre foram essenciais em muitas áreas da investigação e nos últimos anos tornaram-se destaque em mais áreas científicas, crescendo substancialmente em escala e complexidade. Quando a informação é abundante, a própria natureza da investigação começa a mudar, gerando um ciclo de respostas entre investigadores e resultados de investigação que “changes the pace and direction of discovery” (Comissão Europeia: Grupo de peritos de alto nível em dados de investigação, 2010:8). Os Dados Abertos permitem que os investigadores produzam análises e inferências muito além do que os indivíduos e silos de dados isolados podem

¹⁵ Disponível na WWW: <http://hadoop.apache.org/>

¹⁶ Disponível na WWW: http://cordis.europa.eu/project/rcn/206193_en.html

¹⁷ Disponível na WWW: <https://opendataincubator.eu/>

¹⁸ Disponível na WWW: http://cordis.europa.eu/project/rcn/194116_en.html

produzir (Shadbolt, 2016). O International Council for Science (ICSU) considera que quando o setor público é o responsável pelos custos da criação de dados de investigação, é do interesse científico que seja encorajado o máximo e possível uso e reuso dos dados (International Council for Science [ICSU], 2014). A abertura dos dados gerados na investigação científica ou pelas entidades públicas, as grandes fontes de dados na atualidade, constituem uma oportunidade de gerar valor para a sociedade. Segundo Borgman (2015), dados são representações de observações, objetos ou outras entidades usadas como prova de um fenómeno para fins de investigação ou académicos. Os dados tornam-se informação quando são combinados de formas que podem revelar padrões num determinado fenómeno. O conhecimento é gerado pela informação apoiada em dados concretos sobre um fenómeno (Royal Society, 2012).

“In the digital age, data is the raw material on which discoveries are built, and unfettered access to research data” (SPARC). O processo de investigação é dividido em várias etapas, através das quais muitas vezes são produzidos dados primários, estatísticas, cálculos ou anotações, seja em forma bruta ou em forma gráfica, que servem de suporte aos artigos científicos publicados. Contudo, dada a variedade de tipologias e formatos, por norma os dados não são alvo de publicação, correndo o risco de serem perdidos se não forem adequadamente armazenados e preservados. Neste sentido, a abertura de dados é essencial para assegurar a sua preservação e possível reutilização em benefício da sociedade.

Existem várias definições de Dados Abertos, apresentando algumas diferenças entre elas, embora focando o mesmo objetivo. Algumas das definições são emitidas no âmbito da abertura de dados públicos, outras, da abertura de dados de investigação. O portal de dados dos EUA define *Open Data* como os conjuntos de dados de alto valor criados pelo governo federal que possam ser processados por máquina e que não sejam privados ou restritos por razões de segurança (United States General Services Administration). A Royal Society (2012) entende por *Open Data* os dados que devem ser acessíveis, utilizáveis, avaliados ou certificados e inteligíveis. A Open Definition sintetiza que os Dados Abertos devem poder ser “freely used, modified, and shared by anyone for any purpose” (Open Knowledge International [a]). Segundo a Open Knowledge International (OKI) (2012b), “Open Data are the building blocks of open knowledge”. O termo *Open Data* pode ainda referir-se a outros tipos de dados não públicos que sejam acessíveis de forma livre, como os dados dos média sociais (European Data Portal, 2016a).

“Openness is not however enough. Data must be intelligently open” (Science International, 2015:4). O primeiro passo na abertura inteligente dos dados é, segundo a Royal Society (2012:7), “data that underpin a journal article should be made concurrently available in a accessible database”. O mundo académico tem vindo a ser moldado para apoiar a abertura à medida que os benefícios se tornam mais tangíveis e

a pressão de retorno de investimento exercida sobre as entidades financiadoras aumenta, embora nem todas as áreas evoluam com a mesma rapidez (Hahnel & Hook, 2016). A Royal Society (2012) sublinha a necessidade das instituições ligadas à investigação científica gerirem os seus dados de modo proficiente para obterem os benefícios proporcionados pela abertura de dados e entenderem a enorme mudança cultural necessária para atingir este objetivo.

A Denton Declaration foi publicada em 2012 com o intuito de promover boas práticas de gestão de dados de investigação, promover a colaboração, transparência e prestação de contas ao longo das fronteiras organizacionais e disciplinares. Esta declaração estabeleceu treze princípios para promover a cultura de abertura na investigação: estabelecer um padrão para arquivo de dados de modo a facilitar a sua pesquisa e acesso, desenvolver um regime de financiamento para a construção da infraestrutura necessária para a gestão de dados, adotar padrões de metadados, preservar a longo prazo os dados subjacentes às publicações, apoiar os investigadores nas negociações para a abertura dos dados com as editoras e o reconhecer da propriedade intelectual dos investigadores nos dados e em outros produtos de investigação (Denton Declaration: An Open Data Manifesto, 2012, tradução livre).

A gestão de informação e do conhecimento deve ser incluída nas estratégias organizacionais das instituições (Royal Society, 2012). Quando os dados são bem organizados, documentados, preservados e partilhados tornam-se preciosos no avanço científico e no aumento de oportunidades de aprendizagem e inovação (UK Data Archive, 2017). O ciclo de vida dos dados contempla seis fases de gestão incluídas num processo contínuo: a cada reutilização geram-se novos dados, conforme demonstrado na figura seguinte.



Fig. 8. Ciclo de vida dos dados de investigação. Fonte: traduzido de UK Data Archive (2017). Disponível na WWW: <http://www.data-archive.ac.uk/create-manage/life-cycle>

As seis fases do ciclo de vida dos dados comportam diversas ações necessárias à plena execução de cada ciclo:

- Criação: desenho da investigação, criação do plano de gestão de dados, planejar a partilha, localizar dados existente, coletar dados, capturar e/ou criar metadados;
- Processamento: entrada de dados, transcrição, tradução, validação, limpeza, anonimização, descrição, gestão e armazenamento de dados;
- Análise: interpretação e derivação de dados, produção de resultados de investigação, publicação e preparação dos dados para preservação;
- Preservação: migração de dados para o melhor formato, criação de cópia de segurança e armazenamento, criação de metadados e arquivo;
- Fornecer acesso: distribuir e partilhar dados, controlar o acesso, estabelecer a propriedade intelectual, promover os dados;
- Reutilização: acompanhamento da investigação, assumir a revisão da investigação, escrutinar os resultados e aprender e ensinar (UK Data Archive, 2017).

Com o objetivo de contribuir para a adoção de padrões e políticas de Dados Abertos de investigação, algumas organizações lançaram iniciativas e princípios de *Open Data* destinadas à recomendação de boas praticas na publicação de dados de investigação obtidos através de financiamento público. Após o lançamento de uma declaração em 2004, a OECD publicou em 2007 um conjunto de 13 princípios para abertura de dados de investigação provenientes de financiamento público. Em 2010 foram lançados os Panton Principles, um conjunto de quatro princípios delimitadores da abertura de dados na ciência, cujo foco centrou-se no licenciamento de conteúdos para garantir uma clara partilha, distribuição e reutilização de informação pelos investigadores (Panton Principles, 2010, tradução livre). A Royal Society afirma em 2012 que os dados abertos devem ser acessíveis, inteligíveis, avaliáveis e reutilizáveis. Em 2014 foram publicados os Nairobi Data Sharing Principles, um conjunto de dez recomendações destinadas a apoiar a abertura de dados públicos e de investigação nos países em vias de desenvolvimento, por forma a gerar valor económico, social, educacional e potenciar a investigação científica (International Workshop on Open Data for Science and Sustainability in Developing Countries, 2014).

A Open Definition, um projeto da responsabilidade da OKI, fixou os princípios que definem a abertura de dados e conteúdos, assegurando a qualidade e encorajando a compatibilidade de diferentes tipos de conteúdos através da utilização precisa do termo aberto (Open Knowledge International [a], tradução livre). Assim, o termo Dados Abertos engloba dois tipos diferentes de abertura:

- Abertura Técnica: os conteúdos devem ser publicados sob uma licença aberta, ao custo de produção, num formato aberto e que possa ser processado por máquina;
- Abertura Legal: a licença aberta deve permitir o uso, redistribuição, modificação, separação, compilação, propagação, aplicação para qualquer fim ou propósito, sem custos e sem discriminação contra pessoas ou grupos. A mesma licença pode requerer condições de uso como a atribuição de crédito de autoria, manutenção

da integridade do trabalho original, redistribuição sob a mesma licença do original, retenção de notificações de copyright, acesso à fonte original, proibição de implementação de restrições técnicas e implementação de regras de não-agressão com a licença original (Open Knowledge International [a]), tradução livre).

A Science International (2016) é o conjunto das 4 entidades mais representativas a nível da ciência internacional (International Council for Science – ICSU, the InterAcademy Partnership – IAP, The World Academy of Sciences – TWAS and the International Social Science Council – ISSC) que têm por missão representar a comunidade científica global na definição de políticas científicas internacionais. Esta organização lançou 12 princípios para os Dados Abertos, criados com base em outros relatórios da autoria quer dos próprios parceiros da Science International, quer de instituições governamentais ou académicas. Estes princípios assentam em três pilares base, as responsabilidades de cada interveniente, os limites da abertura e as práticas adequadas (Science International, 2015, tradução livre).

Todos estes princípios de abertura de dados de investigação foram sintetizados no Quadro 1, apresentando-se apenas os princípios mencionados em pelo menos duas recomendações de Open Data:

Quadro 1. Princípios de abertura de dados de investigação mais recomendados.

Princípios e recomendações	OECD Princ. (2007)	Panton Princ. (2010)	Royal Soc. (2012)	Nairobi Princ. (2014)	Science Int. (2015)	Open Definition
Abertura / Acesso	✓	✓	✓	✓	✓	✓
Reutilização		✓	✓	✓	✓	✓
Citação de dados				✓	✓	✓
Interoperabilidade	✓			✓	✓	
Sem discriminação		✓		✓		✓
Licenças Abertas		✓			✓	✓
Formato Aberto		✓			✓	✓
Processáveis p/ máquina		✓			✓	✓
Modificação		✓			✓	✓
Publicação imediata				✓	✓	
Qualidade	✓			✓		
Acesso Gratuito				✓		✓
Disponibilidade		✓			✓	
Sustentabilidade	✓			✓		
Qualidade	✓			✓		
Integridade	✓			✓		

Fonte: Elaboração própria.

Analisando em detalhe os princípios de abertura dos dados de investigação é possível retirar algumas ilações sobre quais as recomendações consideradas mais importantes. Verifica-se uma grande variedade de princípios propostos, sendo que

apenas 16 num total de 33 se repetem em 2 ou mais recomendações. A abertura legal e técnica dos dados só é acautelada por metade dos princípios. Os princípios considerados mais importantes são o AA aos dados, seguido da reutilização de dados. A citação de dados, interoperabilidade, acesso sem discriminação, licenças e formatos abertos, a modificação e o fato de os dados serem processáveis por máquina são mencionados apenas em três recomendações, tendo a seu favor o fato de terem sido incluídas nas recomendações mais recentes. Princípios relevantes como a publicação imediata, a qualidade e a integridade dos dados surgem apenas duas vezes listados.

É possível verificar que a OKI (2012b) coloca os dados de investigação ao mesmo nível de importância que outros tipos de dados com potencial para novos usos e aplicações para benefício da sociedade, como os dados culturais, financeiros, estatísticos, ambientais e ainda sobre o clima e transportes.

A OECD (2015a, tradução livre) defende haver quatro tipos de dados científicos:

Dados de observação	<ul style="list-style-type: none"> •provenientes de instrumentos como telescópios, satélites, sensores, inquéritos e outras ferramentas que registem informação sobre fenómenos únicos;
Dados experimentais	<ul style="list-style-type: none"> •obtidos através de máquinas de alto rendimento, ensaios clínicos ou outras experiências controladas;
Dados computacionais	<ul style="list-style-type: none"> •gerados a partir de simulações computacionais;
Coleções de dados de referência	<ul style="list-style-type: none"> •São conjunto de dados agrupados de alto nível de uso pela comunidade científica, recebendo para isso um alto nível de curação

Fig. 9. Tipos de dados científicos. Fonte: traduzido de OECD (2015a)

A explosão de produção de dados que originou o *Big Data* possui um segundo pilar, os Dados Ligados ou *Linked Data*. A combinação destas duas abordagens tem o potencial para criar uma nova era de entendimento científico relativamente aos maiores problemas subjacentes à espécie humana (Science International, 2015). O termo *Linked Data*, cunhado por Tim Berners-Lee, implica que as “relationships among data should be made available” (World Wide Web Consortium [W3C], 2015), ou seja, conjuntos de dados relacionados com um fenómeno em particular e que estão ligados de uma forma lógica podem ser conectados semanticamente dando a possibilidade de um computador poder identificar relações mais profundas entre eles (Science International, 2015, tradução livre). Reitano (2013) sublinha que a informação extraída de conjuntos de dados conectados é a que tem mais impacto na investigação, inovação e crescimento económico. Berners-Lee (2010) propôs um esquema de 5 estrelas para classificar as melhores práticas de publicação de dados na web, que permitam a ligação entre diferentes *datasets*, com cinco níveis de abertura tecnológica:

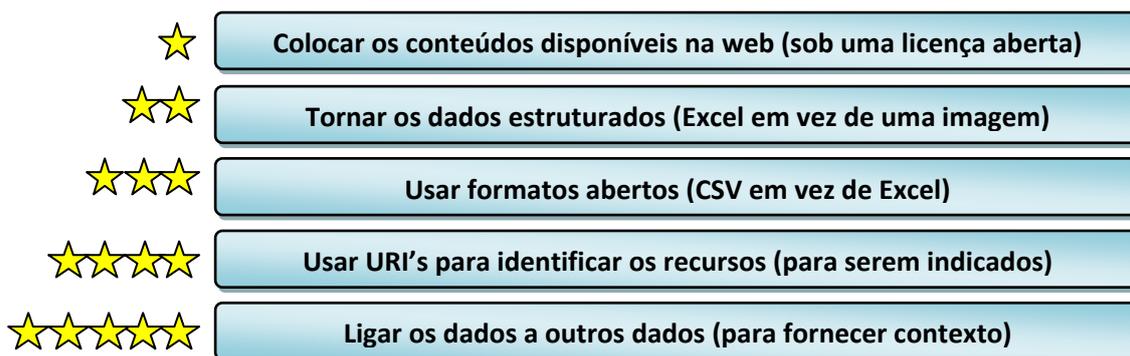


Fig. 10. Esquema de classificação das 5 estrelas para Dados Abertos Ligados. Fonte: traduzido de Berners-Lee (2010). Disponível na WWW: <https://www.w3.org/DesignIssues/LinkedData.html>

O *Linked Data* usa técnicas e conceitos da *World Wide Web* para identificar factos, conceitos, pessoas, lugares, fenómenos ou documentos que contenham atributos em comum (Science International, 2015, tradução livre). A Royal Society (2012:34) afirma que o aparecimento de “siloes sections of the web linked data” deve-se ao uso de vocabulários heterogêneos na descrição semântica dos dados. Neste sentido, a OKI desenvolveu um catálogo de vocabulários de *Linked Data* onde constam um total de 593 distribuídos por vários temas, sendo os mais populares: métodos, metadados e catálogos. Existem 11 vocabulários em português, embora nenhum seja exclusivo, sendo mesmo a língua com maior número de vocabulários a par com o inglês (OKI [a]).

Existem algumas razões para limitar a abertura da ciência, especialmente em redor de dados, que vão além de questões técnicas, envolvendo a comunidade de investigadores e a sociedade em geral. Embora seja mandatado pelas políticas de Dados Abertos que todos os dados de investigação obtidos através de financiamento público devam ser disponibilizados abertamente, o fato é que existem exceções legítimas, como questões de privacidade, segurança ou questões éticas que impedem a total abertura de dados (Concordat on Open Research Data, 2016).

Os desafios tecnológicos não implicam as mesmas preocupações na implementação do AA a dados de investigação, que geram os desafios financeiros, culturais e legais. É necessário “adopt technical and infrastructural solutions that holistically address the above issues”. A atenção é dada a padrões abertos e interoperáveis, pesquisa harmonizada e serviços, identificadores permanentes, promoção da cultura de gestão de dados, reutilização de dados de investigação, soluções técnicas para a segurança e questões legais sobre dados de investigação abertos (RECODE, 2014:6).

Embora se registre progresso, a publicação de Dados Abertos continua a ser um desafio, dado que, a distribuição aberta de dados de investigação não é um comportamento comum, sendo necessário uma mudança cultural (Royal Society, 2012; Pontika et al., 2015; Science International, 2015). A avaliação e revisão por pares dos dados não têm a mesma qualidade quando comparada com as publicações

tradicionais, dado a publicação de dados não ser ainda uma prática comum (European Research Area and Innovation Committee [ERAC], 2016). Hahnel e Hook (2016) apontam a falha na atribuição de incentivos e recompensas tangíveis aos investigadores pelas suas contribuições. A publicação de dados de forma aberta envolve tarefas que consomem muito tempo aos investigadores como a mudança do formato dos dados, manter a informação atualizada, alinhar os dados com as licenças existentes e com os critérios de publicação de informação sensível (Reitano, 2013). A Royal Society (2012) sublinha que uma boa parte dos dados torna-se independente das publicações, separando assim a ideia da evidência.

Para Hahnel e Hook (2016), a abertura de dados científicos não pode ser separada das mudanças no desenvolvimento da investigação e dos receios dos investigadores. Estes receios manifestam-se a dois níveis: estrutural (questões de permissões, qualidade, licenciamento) e cultural (receio de não ser reconhecido ou ser ultrapassado na investigação).

A Science International (2015:6) sublinha que os custos da adaptação ao “data challenge” podem causar problemas aos países menos desenvolvidos, impedindo-os de participar na “global research enterprise”. A mesma organização aponta falhas na replicabilidade de resultados científicos através da citação de três estudos nas áreas de oncologia pré-clínica, psicologia social e economia que obtiveram médias de sucesso bem inferiores a 50%. São apontadas falhas como a falsificação de dados, raciocínios estatísticos inválidos e ausência total ou parcial de dados ou metadados. É também sublinhado que a falta de um padrão de abertura que permita testar a replicabilidade dos resultados científicos pode torna-se uma barreira intransponível no avanço da ciência (Science International, 2015).

O acesso aos dados de investigação possuiu diversas vantagens como o fomento da inovação em termos de educação, saúde, está na base da criação de novos negócios e progresso científico (SPARC; World Wide Web Foundation, 2016), permite a melhoria da qualidade dos resultados de investigação, evita a duplicação de trabalho, acelera a inovação, aumenta a transparência do progresso científico (OECD, 2007; Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016b; ERAC, 2016). A abertura de dados facilita a exploração de novos tópicos não estudados inicialmente, reforça a diversidade de análise e opinião (OECD, 2007), potencia o crescimento económico e a segurança dos resultados da investigação financiada (OECD, 2007; SPARC; Concordat on Open Research Data, 2016; ERAC, 2016). A Royal Society (2012) sublinha que a publicação de dados experimentais e observacionais permite o escrutínio, a replicação de experiência e a reutilização de dados para obtenção de inferências adicionais. A mesma instituição assevera que a abertura de dados permite a identificação de erros, a refinação ou rejeição de teorias, realçando que a análise rigorosa de dados e teoria é o processo mais rigoroso de revisão da qualidade. O

estudo efetuado por Beagrie e Houghton (2014) a três centros de dados do Reino Unido denota um acréscimo significativo ao nível da eficiência da investigação nos utilizadores que usaram os centros de dados e o valor que os utilizadores retiraram dos dados excede o investimento feito na partilha e curação de dados, tendo os centros de dados aumentado exponencialmente o retorno do investimento feito na criação e coleta de dados.

Para Reitano (2013), o processo de abertura de dados pode fomentar a implementação de novos padrões e tecnologias nas organizações e a mudança de comportamentos culturais. O mesmo autor refere que o fenómeno dos Dados Abertos pode dar início ao processo de validação da qualidade dos dados. A Science International (2015) sublinha o investimento nas infraestruturas físicas e tecnológicas não deve ser muito elevado, dado que este não difere muito do exigido para uma base nacional de Ciência Aberta. O desenvolvimento desta e-infraestrutura pode trazer benefícios a vários dos intervenientes da abertura de dados (Comissão Europeia: Grupo de Peritos em Dados Científicos, 2010):

- Cidadãos: beneficiam com os resultados da investigação e encontram as próprias respostas para questões importantes, baseadas em provas reais;
- Financiadores e decisores políticos: tomam decisões mais informadas e têm mais retorno no seu investimento;
- Investigadores: obtém dados e ferramentas com mais facilidade aumentando assim a sua produtividade, a investigação extrapola a sua área de conhecimento, gerando novas inferências e produzindo novas soluções;
- Empresas e indústria: proporcionam uma base industrial e económica para a prosperidade, usam a melhor informação disponível para a sua investigação e desenvolvimento, criam novo conhecimento, mercados, oportunidades e aumentam as oportunidades de troca de conhecimento.

Segundo Borgman (2015:15), o fluxo de informação depende cada vez mais de infraestruturas tecnológicas, sendo que “substantial investments in infrastructure are necessary to acquire, sustain, and exploit research data today, tomorrow and beyond”. As bases de dados científicas têm-se tornado importantes na infraestrutura científica global, sendo indispensáveis no progresso da ciência. De acordo com e-IRG (2016), existem três formas de desenvolver uma infraestrutura de dados de investigação: pelo setor industrial, pelo setor público ou por comunidades especializadas que recebam financiamento para a construção da sua própria infraestrutura. Para melhorar o acesso e uso de *datasets* a e-IRG recomenda:

- Impor as mesmas tecnologias e padrões a todas as disciplinas;
- Atribuir identificadores persistentes para dados, instituições e pessoas;

- Aumentar a pesquisabilidade e descoberta de dados nos repositórios através da criação de registos de *datasets* e repositórios, conectar as publicações aos dados subjacentes e criar uma página de descritiva do *dataset*.

O Open Access Directory (2016) lista 113 repositórios de dados de investigação referentes a 15 áreas do conhecimento. O Registry of Research Data Repositories (re3data) é um serviço que regista repositórios de dados de investigação a nível mundial, fornecendo acesso a 1821 repositórios divididos por quatro grandes áreas científicas. Este serviço é disponibilizado pelo DataCite (2017a), para fornecer acesso e armazenamento permanente de dados aos investigadores, entidades financiadoras, editores e instituições académicas. Portugal conta atualmente com dois repositórios indexados, o KiMoSys e o INTEGRALL, ambos com ligação ao ramo da biologia.

O Projeto do Genoma Humano¹⁹ envolveu uma investigação intensiva à escala global, sendo um exemplo de sucesso da disponibilidade de dados abertos na inovação científica. Com o intuito de facilitar a pesquisa e reutilização de dados, o W3C (2014) desenvolveu um vocabulário específico para dados, o Data Catalog Vocabulary (DCAT), que através da tecnologia RDF²⁰ facilita a interoperabilidade entre diferentes catálogos de dados na web. Utilizar o DCAT para descrever conjunto de dados em catálogos permite aumentar a pesquisabilidade e a extração de metadados a partir de outros catálogos. Salienta-se também a existência do *schema.org*, uma comunidade colaborativa com a missão de criar, manter e promover esquemas para dados estruturados na Internet.

A Data Documentation Initiative (DDI) é um padrão internacional para descrever dados criados por métodos observacionais provenientes das ciências sociais, económicas e da saúde. O DDI é um padrão que pode ser utilizado para documentar as diferentes fases do ciclo de vida dos dados de investigação. Descrever dados utilizando o padrão DDI facilita a compreensão, interpretação e uso por pessoas, *software* e redes de computadores (Data Documentation Initiative [DDI], 2016).

Ainda em formato beta, a OKI oferece um serviço, o *is it open data?*²¹, que permite aos utilizadores inquirirem as entidades detentores de dados sobre a abertura dos seus dados, focando-se essencialmente em dados de investigação.

¹⁹ Este projeto teve como base os Bermuda Principles que asseguraram que todos dados sobre a sequenciação genética, gerados por centros financiados para esse fim, deveriam estar no domínio público com livre acesso. Disponível na WWW:

http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml

²⁰ O Resource Description Framework (RDF) é um modelo padrão desenvolvido pela W3C que permite que dados estruturados e semiestruturados possam ser misturados, expostos e partilhados através de diferentes aplicações. O RDF facilita a união de *datasets* mesmo que usem estes esquemas diferentes. Disponível na WWW: <https://www.w3.org/RDF/>

²¹ Disponível na WWW: <http://www.isitopendata.org/about/>

Verifica-se também o lançamento de revistas científicas específicas para a ciência orientada a dados, como o Data Science Journal²², Scientific Data²³ ou o Geoscience Data Journal, possibilitando a publicação de dados e dos artigos científicos que os utilizam após o serviço de revisão por pares. A F1000 Research²⁴ utiliza o conceito de revisão após a publicação.

Os Dados Abertos têm estado em foco quando se debate a resolução de problemas humanitários. A agenda 2030 definida pelas Nações Unidas (2015) visa a erradicação da pobreza e o desenvolvimento económico, humanitário e social à escala global. Os 17 objetivos de desenvolvimento sustentável contam com a preciosa ajuda da abertura de dados para a monitorização e tomada de decisão. Em 2016 a abertura de dados em casos de emergências públicas foi alvo de forte debate, com a propagação de vírus como Zica e o Ébola na agenda. O debate culminou na Statement on Data Sharing in Public Health Emergencies (2016), um compromisso assinado por 33 organizações a nível global com vista à partilha de dados de investigação e ao fomento do financiamento nestas áreas sensíveis.

Conforme indicado nos princípios de Open Data da Science International (2015), existem diversos intervenientes no processo de abertura de dados sendo que cada um destes intervenientes tem responsabilidades próprias. Como intervenientes diretos são apontados os investigadores que recebem financiamentos públicos, as universidades e institutos de investigação, as bibliotecas institucionais, agências de financiamento e instituições de investigação, governos e sociedade académicas. Existem outras iniciativas/organizações envolvidas na promoção da abertura de dados como a EU, Nações Unidas, OECD, W3C, OKI, SPARC, Royal Society, CODATA, Digital Curation Center (DCC), Research Data Alliance (RDA) ou o ICSU que procuram promover o desenvolvimento de sistemas e procedimentos a nível internacional que assegurem o acesso, sustentabilidade e interoperabilidade. Muitas destas organizações já desenvolveram princípios de abertura e protocolos de dados próprios, como é o exemplo da EU, OECD e da OKI, ou formaram grupos de trabalho para a definição de padrões de gestão, publicação, citação, interoperabilidade e partilha de dados, como a RDA. Outras, como a Universidade de Edimburgo²⁵, criaram cursos de treino em gestão de dados de investigação para oferecer aos seus investigadores esse conhecimento.

Outras atividades e eventos relacionados com a abertura e gestão de dados vão ocorrendo a nível global, como é o caso do Love Your Data²⁶, um evento internacional que teve a sua quinta edição em 2017, com vista a fornecerem conhecimento e ferramentas aos investigadores para que estes possam gerir melhor os seus dados.

²² Disponível na WWW: <http://datascience.codata.org/>

²³ Disponível na WWW: <http://www.nature.com/sdata/>

²⁴ Disponível na WWW: <https://f1000research.com/about>

²⁵ Disponível na WWW: <http://datalib.edina.ac.uk/mantra/>

²⁶ Disponível na WWW: <https://loveyourdata.wordpress.com/lydw-2017/>

1.5 Dados Abertos Governamentais

A abertura dos dados governamentais teve a sua origem em leis de *Freedom of Information* (FOI). Segundo Ubaldi (2013), muitos países da OECD adotaram este tipo de leis para assegurar o direito do acesso público à informação através da publicação de *datasets* que possam ser usados e reutilizados. Os DGA, ou no seu termo mais usual Open Government data (OGD) “is a philosophy- and increasingly a set of policies - that promotes transparency, accountability and value creation by making government data available to all” (OECD, 2016a). Esta filosofia assenta numa nova relação entre cidadãos e Governos, através da abertura dos dados governamentais, onde todas as partes interessadas trabalham em cooperação na procura de novas respostas para os problemas da sociedade num “learning process of changing the culture of public sector administration” (Dietrich, Cherkasov, Susar, 2013:11). Os Governos e as sociedades têm vindo de forma crescente a apreciar a importância do acesso à informação do setor público para sustentar a boa gestão e estimular o crescimento económico alavancado no conhecimento. Esta situação mudou a forma como os governos valorizaram, gerem, usam e partilham informação, sendo apoiada pela atualização das políticas, estruturas e práticas para produzir e usar a informação pública (Ubaldi, 2013, tradução livre). A abertura de dados permite inferir duas questões importantes quanto ao governo. Além de assumir a preparação das entidades públicas para o processo de abertura, também assume que o governo está disposto a efetuar mudanças consideráveis no setor público, devendo atuar como um sistema aberto que interage com o seu ambiente (Jansen, Charalabidis & Zuiderwijk, 2012). O objetivo é, segundo Dietrich (2013:6), passar do “e-gov to we-gov”.

Os DGA estão segmentados em duas partes: os dados abertos e os dados governamentais ou informação do setor público (Ubaldi, 2013; Dietrich, Cherkasov e Susar, 2013), estando os DGA na interseção entre estes dois tipos de dados:

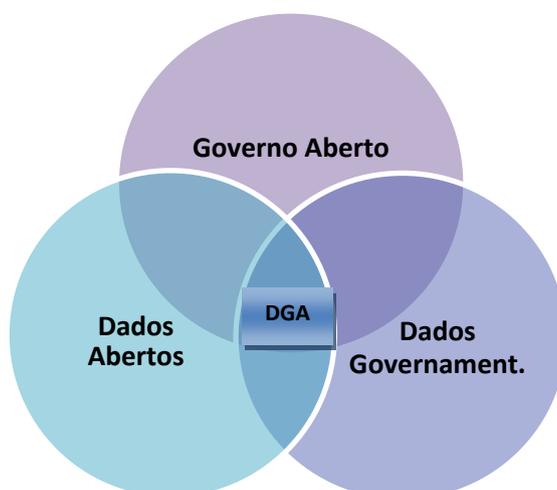


Fig. 11. Pilares dos Dados Governamentais Abertos. Fonte: Dietrich (2013). Disponível na WWW: <http://workspace.unpan.org/sites/Internet/Documents/UNPAN95209.pdf>

Os DGA referem-se à informação coletada, produzida ou paga pelo setor público, guardada em formato que possa ser lido por máquina e tornada acessível de forma livre para reutilização para qualquer propósito, através da publicação sob uma licença aberta que permite a livre reutilização dentro do domínio público e privado (Koski, 2015; Carrara, Oudkerk, Steenbergen, Tinholt, 2016). A Informação do Setor Público ou Public Sector Information (PSI) abrange toda a vasta gama de informação que as entidades públicas coletam, produzem, reproduzem e disseminam em várias áreas de atividade, no cumprimento das suas tarefas institucionais, englobando informação social, económica, geográfica, meteorologia e informação de cariz turístico ou empresarial. A grande diferença entre os DGA e a PSI é que esta última refere-se apenas aos dados detidos pelas entidade públicas, podendo a sua reutilização ter custos. Porém, se esta PSI for disponibilizada sob uma licença aberta, torna-se então DGA (Carrara, et al., 2016a, tradução livre).

A reutilização e distribuição de dados devem ser garantidas sob termos que permitam essas atividades (Ubaldi, 2013). A abertura dos dados governamentais foi apoiada por várias iniciativas com o objetivo de estabelecer padrões e princípios para a disponibilização aberta de dados. Em 2007, o Open Government Working Group (OGWG) definiu oito princípios sobre a abertura de Dados Governamentais (Tauberer, 2014). Em 2010 foi a vez da Sunlight Foundation publicar 10 princípios para abertura da informação governamental. Em 2012 foram lançadas três iniciativas de abertura de dados. O Reino Unido lançou o Open Data White Paper que incluiu um conjunto de 14 princípios para a abertura de dados públicos e assevera que o Reino Unido não pretende implementar legislação de forma rápida, mas sim reconhecer e clarificar medidas iniciais que assegurem a partilha de dados (Cabinet Office, 2012). O Banco Mundial publicou 13 princípios de abertura de dados integrados no seu Open Government Data Toolkit. A OKI (2012a, tradução livre) identificou as componentes chave da abertura de dados públicos:

- Disponibilidade e acesso – os dados devem estar completos e a um preço não superior ao seu custo de reprodução, com disponibilidade de acesso pela Internet. Devem estar disponíveis de forma conveniente e modificável;
- Reutilização e redistribuição – os dados devem ser disponibilizadas sob licenças que permitam a reutilização, redistribuição e mistura com outros *datasets*. Deve ser assegurado que os dados são passíveis de serem processados por máquina;
- Participação universal – os dados devem ser acessíveis a todos, para livre reutilização e redistribuição, sem discriminar nenhum campo do conhecimento, pessoas ou grupos. Não é permitido a inclusão de restrições adjacentes aos dados.

Ubaldi (2013) publicou onze princípios de abertura de dados baseados em dois vetores principais: disponibilidade e acessibilidade e a reutilização e redistribuição. Nos

EUA foram lançados 7 princípios para a abertura de dados através do Memorando “Open Data Policy - Managing Information as an Asset” (Executive Office of the President: Office of Management and Budget, 2013). O Anti-corruption Working Group do G20²⁷ promoveu, em 2014, a abertura de dados como um dos pilares essenciais da transparência do setor público, estabelecendo seis princípios de abertura de dados para apoiar a luta contra a corrupção (G20’s Anti-corruption Working Group, 2014).

O G8²⁸ iniciou um debate sobre a transparência, inovação e prestação de contas governamentais, culminando na criação do G8 Open Data Charter (Cabinet Office, 2013), “which outlined a set of five core open data principles” (International Open Data Charter, 2015), ou seja, “a summary of visions and principles for creating a transparent Government, the opening up of data and its quality and quantity” (Carrara, et al., 2016a:39). O G8 Open Data Charter criou uma lista de 14 categorias de dados a serem publicados de forma aberta: empresas e negócios, crimes e justiça, observação da terra, educação, energia e ambiente, finanças e contratos, geoespacial, desenvolvimento global, saúde, democracia e contabilidade governativa, ciência e investigação, estatísticas, mobilidade e bem-estar social e por fim, transportes e infraestruturas (Cabinet Office, 2013). Porém, o sentimento geral era que os princípios deveriam ser redefinidos e melhorados de forma a apoiar a adoção global dos Dados Abertos (International Open Data Charter, 2015, tradução livre). Assim, em 2015 um grupo de vários países e organizações internacionais desenvolveu a International Open Data Charter, uma “Global Multi-Stakeholder Action Network with two types of leading members: Stewards and Lead Stewards”, tendo como missão fomentar uma maior colaboração na adoção e implementação de princípios, padrões e boas práticas de Dados Abertos em todo o mundo, com a adesão de 41 governos (International Open Data Charter, 2015). Tim Berners-Lee, numa opinião citada no próprio sítio web da International Open Data Charter (2015), vê nesta rede o “potential to accelerate progress by placing actionable data in the hands of people”. Todos os princípios e recomendações de abertura de dados públicos foram sintetizados, apresentando-se no Quadro 2 apenas os princípios mencionados em duas iniciativas de *Open Data*:

²⁷ O G20 é um grupo formado pelos 19 países mais fortes economicamente, mais a União Europeia, com o objetivo de obter cooperação internacional em questões financeiras e económicas. Disponível na WWW: https://www.g20.org/Webs/G20/EN/G20/FAQs/faq_node.html

²⁸ O G8 é um grupo constituído pelas economias mais avançadas do mundo que se reúnem para debater questões económicas e políticas a nível global. Este grupo é formado pelos EUA, Itália, Alemanha, Reino Unido, França, Japão, Canadá e Rússia, embora esta última tenha sido expulsa em 2014 após a anexação da Crimeia. A EU também está representada neste grupo, embora, não seja membro oficial.

Quadro 2. Princípios de abertura mais recomendados pelas iniciativas de Dados Governamentais Abertos.

Princípios e recomendações	OGWG (2007)	Sun. Fund. (2010)	OKI (2012)	World Bank (2012)	White Paper (2012)	Mem. EUA (2013)	Ubaldi (2013)	G20 (2014)	IODC (2015)
Abertura / Acesso	✓	✓	✓	✓	✓	✓	✓	✓	✓
Publicação imediata	✓	✓		✓	✓	✓	✓	✓	✓
Reutilização			✓	✓	✓	✓	✓	✓	
Completos	✓	✓	✓	✓		✓	✓		
Sem discriminação	✓	✓	✓		✓		✓		
Processáveis p/ máquina	✓	✓	✓		✓		✓		
Licenças Abertas	✓	✓	✓		✓				
Comparabilidade			✓	✓				✓	✓
Formato Aberto		✓	✓		✓		✓		
Disponibilidade			✓	✓		✓	✓		
Abertura p/ defeito			✓					✓	✓
Interoperabilidade			✓					✓	✓
Publicação de metadados		✓			✓	✓			
Gestão pós publicação		✓		✓		✓			
Custo não superior ao da produção		✓	✓						
Desenvolvimento e inovação inclusivos								✓	✓
Melhoria governação e envolvimento cidadãos								✓	✓
Linked Data					✓		✓		
Modificação			✓				✓		

Fonte: Elaboração própria.

Analisando em detalhe os princípios de abertura dos dados públicos é possível inferir quais as recomendações consideradas mais importantes. Tal como nos dados de investigação verifica-se uma grande variedade de princípios propostos, sendo que apenas 19 num total de 36 se repetem em 2 ou mais recomendações. Embora se registre uma melhoria significativa na abertura legal e técnica relativamente aos dados de investigação, também os princípios propostos para os dados públicos não abarcam a totalidade desta dupla abertura. O acesso, a publicação imediata, a reutilização e a disponibilização de dados de forma completa, sem discriminação e de forma a serem processados por máquina são os princípios que reúnem mais consenso entre as recomendações analisadas. As licenças abertas, a comparabilidade de dados, os formatos abertos, a disponibilidade, a abertura por defeito de dados, a interoperabilidade, publicação de metadados e a gestão de dados após a sua publicação são mencionados em menos de metade das recomendações analisadas, apesar de ser indiscutível a sua importância no processo de abertura de dados.

Segundo Gonzalez-Zapata e Heeks (2015), podemos sintetizar os DGA em quatro perspectivas sumariadas na figura seguinte:

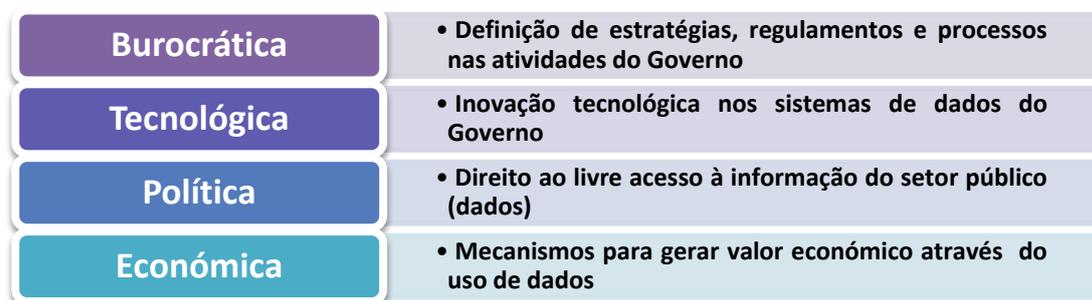


Fig. 12. Perspetivas da abertura dos dados governamentais. Fonte: Gonzalez-Zapata e Heeks (2015)

O movimento dos DGA assenta em dois importantes marcos: o lançamento do sítio data.gov nos EUA em 2009 e do data.gov.uk no Reino Unido em 2010, tendo-se posteriormente estendido a outros países e organizações (Janssen, 2012). Neste segmento foi fundada em 2011 a Open Government Partnership (OGP) com o objetivo de “provide an international platform for domestic reformers committed to making their governments more open, accountable, and responsive to citizens” (Open Government Partnership [OGP], 2015). Esta organização lançou em Setembro de 2011 a Open Government Declaration, subscrita por 70 países, e que assenta em quatro eixos estruturantes: Transparência da informação sobre as atividades governativas; apoio à participação cívica; implementação dos mais altos padrões da integridade profissional; aumento do acesso a novas tecnologias de abertura e da prestação de contas (OGP, 2011). Em Dezembro de 2016 realizou-se a quarta cimeira da OGP, culminando com a divulgação da Paris Declaration for Open Government Partnership que pretende encorajar os Governos e as organizações da sociedade civil a inscreverem-se em ações coletivas para promover a abertura dos dados governamentais (OGP, 2016).

O fenómeno dos Dados Abertos é coadjuvado por dois sistemas de avaliação da abertura dos dados provenientes do setor público que auxiliam as entidades públicas a assumir a abertura e ligação de dados (Colpaert, Sarah, Peter, Mannens & Van de Walle, 2013). Estes dois sistemas atuam em perspetivas diferentes, segundo Colpaert, *et al.* (2013): o 5 Stars Linked Open Data de Berners-Lee (2010) fornece uma perspetiva tecnológica enquanto o 5 Stars Open Data Engagement de Davies (2012) tem um âmbito social através do envolvimento da sociedade. O esquema de Davies (2012) é um esquema de classificação por estrelas, baseado no esquema criado Tim Berners-Lee, para sublinhar os níveis de interação e colaboração proporcionados pela publicação de dados, ou seja, os “key steps that open data initiatives can take to engage with data users”.

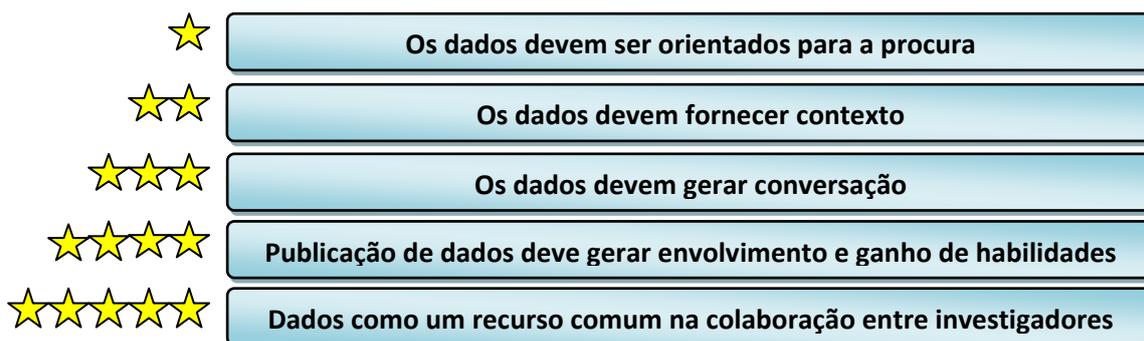


Fig. 13. Esquema de classificaço das 5 estrelas para as iniciativas de abertura de dados que envolvem os utilizadores. Fonte: Davies (2012). Disponvel na WWW: <http://www.timdavies.org.uk/2012/01/21/5-stars-of-open-data-engagement/>

No cumprimento das suas obrigaço, as entidades governamentais produzem, coletam e gerem uma vasta quantidade de dados, tendo elevado estes à categoria dos bens púlicos mais valiosos. Para gerar valor “data must be reusable, but more importantly, actually used” (Open Data for Development [OD4D], 2016:8). Para a OECD (2015a), permitir o acesso e a reutilizaço dos dados governamentais tem um potencial significativo no aumento da eficiêcia e transparêcia do setor púlico, suportando atividades de ciêcia cidadã com os dados governamentais, aumentando assim o seu valor econmico e social. Neste sentido, concursos, desafios e *hackathons*²⁹ podem tornar-se úteis se ajudarem a focar os cidads “on high priority public problems” (Dawes, Vidiarova & Parkhimovich, 2016:10). Para gerar valor é necessrio publicar os dados, sendo esta fase precedida uma preparaço tcnica de coleta, preparaço, publicaço e gesto de dados (Carrara, et al., 2016a):

- Criaço de uma infraestrutur de gesto de dados;
- Extraço, transformaço (em formato *open source* e *linked data*) e publicaço;
- Escolha dos canais de publicaço;
- Implementaço da funço de pesquisa no canal de publicaço;
- Definiço de domnio, hospedagem, funcionalidades do canal de publicaço, incluindo a gesto desta ferramenta e dos servios que nela forem incorporados.

Segundo Ubaldi (2013, traduço livre), podem ser identificadas quatro grandes fases na cadeia de valor dos Dados Governamentais Abertos:

- Geraço de dados: cobre a criaço de dados púlicos;
- Coleta, agregaço e processamento de dados: os dados governamentais precisam de ser agregados, ligados e manipulados de forma a adicionar valor para a maior parte dos utilizadores;

²⁹ Uma *hackathon* de dados abertos é um evento que reúne programadores e outros profissionais do ramo do desenvolvimento de *software* para reutilizar dados de forma a produzir aplicaço de *software* que sejam a base de produtos e servios úteis e com valor. Servem tbm para promover o trabalho de abertura de dados e os *datasets* disponveis (AMA, 2016).

- Distribuição e fornecimento de dados: os dados precisam de ser distribuídos aos seus potenciais utilizadores para permitir o acesso e reutilização;
- Uso final de dados: os dados governamentais têm de ser reutilizados por diferentes pessoas para sustentar a criação de valor público.

Perceber os diferentes valores associados aos dados governamentais é essencial para identificar que viabilizadores e tipos de dados devem ser tornados prioritários para atingir o objetivo específico de criação de valor (Ubaldi, 2013). Podem ser identificados três tipos de valor associados aos DGA (Ubaldi, 2013; OECD, 2015a):

- Valor económico: crescimento e competitividade na economia, fomentando a inovação, eficiência e eficácia nos serviços governamentais;
- Valor Social: promove empreendedorismo dos cidadãos, a participação social e o envolvimento público nas decisões políticas e no fornecimento de serviços;
- Governança pública: Prestação de contas, transparência, responsividade e controlo democrático.

A abertura dos dados governamentais é um domínio emergente que enfrenta restrições de natureza complexa. Janssen, Charalabidis e Zuiderwijk (2012) sintetizaram as barreiras em seis categorias: institucional, complexidade da tarefa, uso e participação, legislação, qualidade da informação e questões técnicas. Reitano (2013) defende que os governos devem modificar as práticas e procedimentos internos de gestão de informação, pois a reutilização da informação está dependente da forma em como ela é gerida e publicada. O desempenho dos programas de DGA tendem apenas para o registo de dados estatísticos e poucos optam por analisar as aplicações comerciais e os avanços proporcionados pela abertura dos dados governamentais (Dawes, Vidiasova & Parkhimovich, 2016). Existem outros pontos sensíveis como o pagamento dos custos da coleta e processamento de dados públicos, quais os incentivos concedidos às entidades públicas para estas manterem e atualizarem os seus dados, que conjunto de dados devem ser prioritários para maximizar a criação de valor público e, sobretudo, a necessidade de desenvolver uma estrutura de análise entre custos e benefícios que possa demonstrar as vantagens económico-sociais dos DGA (Ubaldi, 2013; OECD, 2013). Contudo, segundo a OECD (2015a), para as partes interessadas estes dados continuam muitas vezes inacessíveis, quer em termos de acesso, quer em termos de custo. Isto verifica-se porque os dados enfrentam alguns entraves à sua publicação: as organizações e governos não consideram os dados interessantes para reutilização ou não conseguem entender os benefícios da abertura, pelo receio de terem de responder a um elevado número de pedidos de dados ou que os utilizadores retirem conclusões superficiais, que os dados possam não ter um nível de curaçao satisfatório e que isso manche a reputação da organização, e por fim, a preocupação com os custos da conversão de dados para formatos padrão (Carrara, et al., 2016a, tradução livre). Existem ainda riscos associados à abertura dos dados

governamentais como o direito à privacidade e ao anonimato que resultam em restrições à disponibilização desta informação (Koski, 2015).

Estabelecer um enquadramento legal é essencial para alavancar a abertura dos dados governamentais. Dawes, Vidiasova e Parkhimovich (2016) defendem que o sucesso dos programas de DGA depende da definição padrões técnicos e de qualidade e de um enquadramento legal que defina responsabilidades e limitações. Ainda assim, as iniciativas de DGA podem enfrentar problemas com a implementação, desafios técnicos ou atrasos administrativos. Restrições técnicas e financeiras podem limitar a acessibilidade e reutilização dos dados, sendo necessário enfrentar desafios relacionados com políticas, tecnologias, finanças, organização e cultura para criar um ecossistema e construir um modelo de negócios sustentável para as iniciativas de DGA (Ubaldi, 2013, tradução livre). Koski (2015) realça que uma vez abertos, deve haver um esforço para manter a qualidade dos dados e também a sua atualização. Por outro lado, sem os conhecimentos e habilidades necessárias para interpretar a vasta gama de conjuntos de dados, os cidadãos podem não conseguir criar valor a partir desses dados (Janssen, 2012; Koski, 2015). Segundo Carrara, Fischer e Steenbergen (2015), são várias as capacidades necessárias para lidar com a complexidade dos dados. As autoras defendem uma junção entre *soft skills* e *hard skills* como a colaboração, comunicação, resolução de problemas e as competências técnicas, estatísticas, analíticas e domínio da área do conhecimento (2015:12). São apontados ainda outros desafios como a constante evolução das políticas e estratégias de DGA, a delegação da produção e administração dos dados governamentais a nível regional e a falta de definições comuns à escala internacional (OECD, 2013).

Dado o contexto de urgência em que muitos dos portais foram criados, “as part of a politically driven Open Data Initiative, without feasibility studies, business case, strategy or user research”, os portais enfrentam agora uma série de desafios e problemas (Sasse, et al., 2017:7). São mencionados problemas como a falta de financiamento, falta de qualidade dos dados, a falta de apoio político na publicação de dados relevantes e desafios na reutilização de dados (Sasse, et al., 2017).

Um bom entendimento dos benefícios da abertura dos dados governamentais pode acelerar o compromisso com as iniciativas de DGA (Carrara, et al., 2016a). Os benefícios da abertura de dados podem, segundo Janssen, Charalabidis e Zuiderwijk (2012), ser contextualizados em três quadrantes: político e social, económico, operacional e técnico. A partilha dos dados governamentais com o público aumenta a transparência e o reconhecimento público das atividades do governo (OECD, 2013), dado que, os DGA são um recurso estratégico que pode beneficiar vários intervenientes (OECD, 2013; Ubaldi, 2013; Dietrich, Cherkasov, Susar, 2013; Koski, 2015; Carrara, et al., 2016a, The World Bank, 2016):

- Governo: melhora a tomada de decisões e a alocação de recursos, de forma a promover a eficiência total das operações do governo e fornecer serviços públicos mais eficientes, efetivos, inovadores e personalizados, permitindo uma redução de custos administrativos. Aumenta a transparência, reforçando a prestação de contas governamentais e a redução de erros;
- Cidadãos: permite participação pública e o envolvimento social no desenho de respostas às necessidades públicas e científicas (ciência cidadã);
- Sociedade Civil: aumenta a transparência e qualidade dos serviços do Governo;
- Setor Público: pode estimular um mercado competitivo nos serviços públicos;
- Setor Privado: lançamento de serviços inovadores e com valor acrescentado.

O Working Group on Open Government Data da OKI elegeu três vantagens na abertura dos dados governamentais: aumento da transparência, criação de valor económico-social e participação dos cidadãos na governação. Dietrich, Cherkasov e Susar (2013) sublinham o papel chave da interoperabilidade na obtenção dos benefícios da abertura, na medida em que uma definição comum de abertura irá assegurar a capacidade dos sistemas e organizações cooperarem.

O impacto da abertura dos dados governamentais pode ser medido em três dimensões: económica, política e social (Granickas, 2013; Koski, 2015):

- Económica: aumento de receitas provenientes de taxas, redução dos custos de transação de dados, criação de bens e serviços (Granickas, 2013), aumento da eficiência e eficácia do serviço público, criação de novas empresas (Granickas, 2013; Koski, 2015);
- Política: aumento da transparência política e da prestação de contas (Granickas, 2013; Koski, 2015);
- Social: aumento da proteção ao direito à informação (Granickas, 2013), aumento da inclusão social e da capacitação de grupos e/ou indivíduos, aumento dos níveis de envolvimento e participação cívica (Granickas, 2013; Koski, 2015) e melhorias ao nível educação, saúde, ambiente ou desenvolvimento sustentável (Koski, 2015).

Koski (2015:3) defende ser necessário desenvolver um modelo de avaliação e monitorização seguido da coleta sistemática de dados para se poder medir com exatidão o impacto dos Dados Abertos. A autora sublinha a necessidade de explorar o conhecimento sobre o impacto da abertura de dados, bem como os métodos de investigação e indicadores que podem ser usados para estudar os “socio-economics impacts of open data, which types of usable data are already being collected and what are the additional data needs”.

Para Janssen (2012), o movimento dos DGA promoveu uma maior disponibilidade de informação pública, embora não tenha necessariamente melhorado o acesso à informação. De acordo com a mesma autora, os DAG estão em voga, com os

programadores e *hackers* a desenvolverem aplicações através dos dados detidos pelo setor público e os ativistas a descobrir o valor dos dados governamentais na defesa das suas causas. Se as cidades utilizarem os dados que produzem em proveito próprio, publicando-os nos seus portais com as licenças adequadas de reutilização, podem estimular a criação de novos serviços e transformar-se em *smart cities* (Garcia, 2015), reduzindo custos e consumo de recursos e melhorando o contacto entre cidadãos e governo (Carrara, Engbers, Nieuwenhuis & Steenbergen, 2016).

No âmbito da *digital transformation* ocorrida nas últimas décadas, diversas alterações ocorreram na forma como as organizações produzem, vendem e disponibilizam os seus produtos e também na forma em como os funcionários trabalham e comunicam. Segundo Carrara, Fisher, Oudkerk, Steenbergen e Tinholt (2015), os Dados Abertos podem ser reutilizados por organizações de diferentes setores da indústria: cultura e turismo, educação e investigação, energia e utilidades, finanças e seguros, saúde, fabrico de produtos, comunicação e entretenimento, público, imobiliário, retalho, telecomunicações, transportes e tecnologias de informação. Davies (2014) afirma que no caso dos EUA e do Reino Unido a ênfase do discurso público passou da melhoria da transparência governativa para o discurso das vantagens económicas dos Dados Abertos, como consequência da crise económica e da maior maturação das políticas de abertura de dados. Na EU a linha de pensamento tem evoluído no mesmo sentido, com o estudo económico feito pelo European Data Portal a indicar um impacto económico direto da abertura de dados em 2016 de 55.3 biliões de euros e em 2020 de 75.7 biliões de euros, com uma projeção de perto de 100 mil postos de trabalho envolvidos. O mesmo estudo indica que o mercado total da abertura de dados, contabilizando os impactos diretos e indiretos subirá dos 193-209 biliões de euros em 2016 para os 265-286 biliões de euros em 2020 (Carrara, Chan, Fischer & Steenbergen, 2015).

Os portais de Dados Abertos são uma parte essencial da infraestrutura ao permitirem o acesso e reutilização de dados, oferecendo um ponto de acesso único aos utilizadores aos dados governamentais como os catálogos e portais de dados centralizados. Veja-se, por exemplo, os 304 portais incluídos no catálogo do portal de dados governamentais dos EUA, o data.gov, que além dos portais de dados de cidades e estados norte-americanos (88), engloba ainda 52 países e 164 regiões internacionais (United States General Services Administration). O Open Data Soft (2017) lista já mais de 2600 portais de dados oriundos de 194 países, indicando apenas 6 portais de Dados Abertos em Portugal.

Porém, de acordo com Colpaert, *et al.* (2013) não existe nenhum guia que determine as funcionalidades e possibilidades dos portais de dados. Os autores vêm assim propor um esquema de classificação para portais de Dados Abertos com o

objetivo de impulsionarem as entidades públicas a criarem o seu próprio portal de dados ou a criarem novas metas para os já existentes.

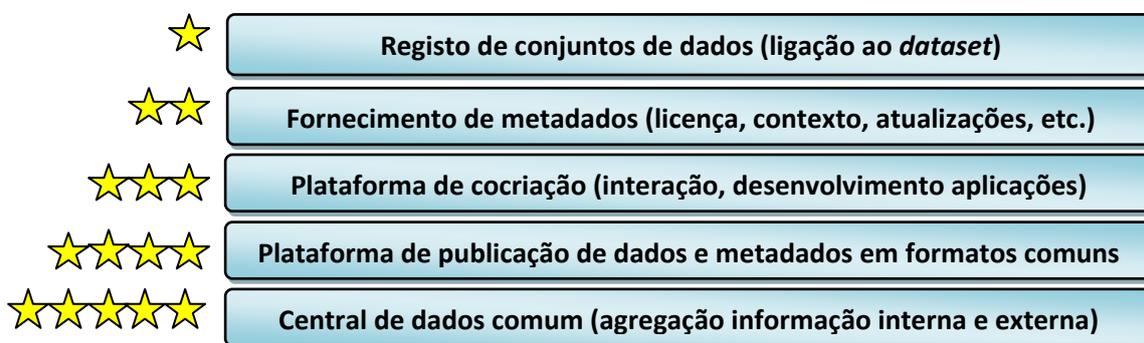


Fig. 14. Esquema de classificação das 5 estrelas para os portais de dados. Fonte: Colpaert, et al. (2013). Disponível na WWW: <http://pieter.pm/5stardataportals/5starsofopendataportals.pdf>

O relatório feito pelo Open Data Institute (ODI) para o European Data Portal defende que os portais de dados devem continuar a responder aos desafios utilizando para isso modelos sustentáveis em quatro áreas chave: gestão, finanças, arquitetura e operações. Devem também usar métricas que permitam avaliar e melhorar a qualidade e usabilidade dos dados, sendo que a medição de impacto da reutilização de Dados Abertos também deve ser alvo de financiamento contínuo (Sasse, et al., 2017).

A OKI desenvolveu um sistema de gestão de dados suportado por uma plataforma aberta, o CKAN, oferecendo soluções que agilizam a publicação, partilha, pesquisa e uso dos dados, sendo utilizado por 146 portais de dados a nível mundial (CKAN). Usar esta ferramenta para construir um catálogo permite às entidades, governos e organizações publicar dados e disponibiliza-los de forma aberta, facilitando a pesquisa e a visualização da informação. A plataforma DataHub oferece acesso a muitos dos serviços do CKAN, permitindo o *upload*, pesquisa, registo de *datasets* publicados, criação e gestão de grupos de *datasets* e receber atualizações de conjuntos de dados em que haja interesse. O DataHub aloja 35 *datasets* relacionados com a palavra-chave “Portugal”, sendo 18 desses conjuntos de dados diretamente coletados e disponibilizados por entidades portuguesas. Os restantes 17 *datasets* estão englobados em publicações que abrangem vários países (Open Knowledge International [d]). As entidades públicas também podem desenvolver portais de dados, contratando para isso serviços de plataformas como o Socrata³⁰, OpenDataSoft³¹ ou o DataMarket³² que facilitam a publicação, a partilha e a reutilização de dados públicos.

³⁰ Disponível na WWW: <https://socrata.com/>

³¹ Disponível na WWW: <https://www.opendatasoft.pt/solucao-de-dados-abertos/>

³² Disponível na WWW: <http://www.qlik.com/us/products/qlik-data-market>

Em termos de licenciamento de dados, embora as licenças Creative Commons possam ser aplicadas aos dados, foram criadas licenças abertas específicas para dados sob a tutela da OKI, as Open Data Commons (ODC):

- ODC Attribution License (ODC-By): Compatível com a licença CC BY, permite a cópia, distribuição, modificação de dados ou bases de dados com a devida menção de atribuição de autoria, devendo ser mantida para trabalhos derivados;
- ODC Open Database License (ODC-ODbL): Segue os mesmos princípios da licença CC BY-SA, aplicando-se no entanto apenas a dados. Permite a cópia, distribuição, modificação de dados ou bases de dados com a devida menção de atribuição de autoria, devendo os trabalhos derivados aplicar a mesma licença e serem redistribuídos sem restrições;
- Public Domain Dedication and License (PDDL): permite a cópia, distribuição, modificação de dados ou bases de dados sem qualquer tipo de restrições ou atribuição de crédito, uma vez que os dados passam para o domínio público. É compatível com a licença CC0 (Open Knowledge International [d], tradução livre).

Existem ainda 2 países que definiram a sua própria licença aberta de forma a garantir a abertura da PSI. O Reino Unido desenvolveu a *Open Government Licence* (OGL), que sendo compatível com as licenças CC e ODC autoriza o uso, distribuição, publicação, adaptação e exploração comercial e não comercial sob o requerimento de menção da fonte e da proviência da informação (The National Archives). Também a França desenvolveu a sua *Licence Ouverte* que avaliza a reprodução, uso, distribuição, publicação, adaptação e exploração comercial, requerendo apenas o reconhecimento da fonte e da proviência da informação. Tal como a licença britânica, a *Licence Ouverte* é compatível com as licenças CC e ODC e ainda com a própria OGL (Etalab, 2011).

A publicação dos DGA envolve quatro intervenientes diferentes (Carrara, et al., 2016a, tradução livre):

- Decisores políticos: normalmente os responsáveis pela validação de estratégias e iniciativas de *Open Data*;
- Gestores de Dados: são por norma os responsáveis pela coleta, partilha de dados, *design* e implementação da estratégia de Open Data na qualidade de especialistas;
- Programadores: responsáveis pela implementação dos requisitos técnicos das estratégias de Dados Abertos;
- Contribuidores: têm um papel ativo na coleta, preparação, publicação e manutenção dos dados.

Gonzalez-Zapata e Heeks (2015) acrescentam outros intervenientes como os ativistas da sociedade civil, as organizações internacionais, os fornecedores de tecnologias de informação e comunicação e os académicos. As entidades e organizações internacionais envolvidas no esforço de abertura dos dados

governamentais têm desenvolvido esforços para promover a abertura de dados governamentais. O Open Data Institute (2012) foi criado com o objetivo de ajudar instituições públicas, organizações comerciais e não comerciais a “identify and address how the web of data will impact their businesses and their sectors”. Neste sentido, a ODI tem promovido algumas iniciativas como o Open Data Certificate³³ que, através de um sistema de questionários consegue demonstrar se os dados foram publicados de forma sustentável e reutilizável, certificando assim a qualidade e a confiança nos mesmos. A ODI desenvolveu também o Open Data Maturity Model para avaliar o quanto as organizações publicam e consomem Dados Abertos, identificando os pontos de melhoria. O modelo baseia-se numa escala de cinco temas: processos de gestão de dados, conhecimento e habilidades, envolvimento e apoio ao cliente, investimento e performance financeira e por fim, visão estratégica. Estes cinco temas têm cinco níveis de progressão, possuindo uma aplicação³⁴ em linha que facilita a avaliação (Dodds & Newman, 2015). A ODI dedica-se ainda à elaboração de estudos de caso envolvendo iniciativas de Open Data, além de possuir guias de orientação para a abertura de dados e cursos de aprendizagem com o mesmo objetivo.

O Banco Mundial lançou o Open Government Data Toolkit com o objetivo de ajudar governos, bancários e utilizadores a entender os conceitos básicos da abertura de dados e a planear e implementar um programa de abertura dos dados governamentais. Falamos de ferramentas de orientação para a implementação de políticas de dados abertos, desde a definição de Open Data e exemplos do seu uso, passando pelo licenciamento e qualidade dos dados e finalizando com assistência técnica e apoio na procura de financiamento. Uma das ferramentas disponibilizadas é o Open Data Readiness Assessment (ODRA) que pode ser usada para conduzir ações de análise da preparação de um governo ou instituição para avaliar, desenhar e implementar uma iniciativa de Dados Abertos (The World Bank, 2016).

Outra iniciativa, esta com a chancela da Open Data for Development Network (OD4D[a]), “a global network of leaders in the Open Data community, working together to develop open data solutions around the world”, desenvolveu uma base de dados pública de organizações que usam DGA ao nível global. O Open Data Impact Map permite aos governos, organizações internacionais e investigadores ter a perceção da procura existente de Dados Abertos. O mapa inclui organizações que usam DGA para advogar algo, para desenvolver produtos e serviços, aprimorar operações, prestar informações sobre estratégias ou efetuar investigação (Open Data for Development Network (OD4D[b], tradução livre). Em fevereiro de 2017 encontram-se listadas 1765 organizações oriundas de 96 países, concentrados sobretudo na América do Norte, Europa e Ásia Central, sendo que a maioria das entidades está relacionada com o setor das tecnologias de informação, geoespacial, negócios,

³³ Disponível na WWW: <https://certificates.theodi.org/en/>

³⁴ Disponível na WWW: <http://pathway.theodi.org/>

consultadoria, investigação e com governabilidade. Portugal conta com 4 organizações listadas: Fundação Francisco Manuel dos Santos³⁵, Transparência Hackday³⁶, e Open Knowledge Portugal no setor da governabilidade e a Manufatura Independente³⁷ no setor dos média e comunicações.

A OKI está ativamente envolvida em diversos projetos que visam criar condições e potenciar a abertura de dados. Além da mencionada Open Definition, a OKI posiciona-se como coordenadora de outros projetos semelhantes: um sítio web sobre Dados Governamentais Abertos, um guia com enfoque nos aspetos técnicos, legais e sociais dos Dados Abertos, um catálogo de portais de Dados Abertos a nível global, um índice que mede a abertura das políticas de *Open Data* e ainda introduziu censos sobre a abertura de dados. A OKI oferece ainda um portal sobre dados financeiros³⁸, um portal de *e-learning* sobre literacia de dados³⁹, um portal sobre gastos públicos⁴⁰, uma base de dados de ensaios clínicos⁴¹ e o Frictionless Data⁴² que disponibiliza um Data Package Standard para facilitar o agrupamento e descrição de dados.

A EU tem assumido uma dupla função na abertura dos dados públicos, verificando-se a sua intervenção na elaboração de leis de apoio aos DGA e na promoção da criação de infraestruturas de suporte. A OECD desenvolveu um Observatório da PSI que coleta e analisa exemplos e partilha experiências de inovação do setor público com o intuito de fornecer conselhos práticos aos países na criação de inovação. O Observatório permite a partilha, discussão e cooperação na criação de soluções (OECD, 2016b, tradução livre).

A disponibilidade e acessibilidade de dados é um dos pilares da plena afirmação da Ciência Aberta e neste âmbito devem ser definidas políticas de Dados Abertos que mandatem o livre acesso a este tipo de informação científica e administrativa.

2 As políticas de Dados Abertos

O processo de desenvolvimento de políticas envolve, tradicionalmente, a investigação, análise, consulta e síntese de informação para produzir recomendações. A abertura de dados sem restrições de acesso deve ser garantida por uma política de Dados Abertos, pois “Open Data is political” (World Wide Web Foundation, 2016:38). “Institutionalizing open data within larger sectoral initiatives requires ample political

³⁵ Disponível na WWW: <http://www.pordata.pt/>

³⁶ Disponível na WWW: <http://www.transparenciahackday.org/>

³⁷ Disponível na WWW: <http://manufacturaindependente.org/>

³⁸ Disponível na WWW: <http://next.openspending.org/>

³⁹ Disponível na WWW: <http://schoolofdata.org/>

⁴⁰ Disponível na WWW: <http://openbudgets.eu/>

⁴¹ Disponível na WWW: <http://opentrials.net/>

⁴² Disponível na WWW: <http://frictionlessdata.io/>

will” (OD4D, 2016). O fenômeno dos Dados Abertos é também técnico e social, dada a interação entre o meio local, político, social, legal e técnico e as tecnologias que produzem resultados (Davies, 2014; Pasqueto, Sands & Borgman, 2015).

Da revisão da literatura destaca-se a existência de políticas diferenciadas para a abertura de dados governamentais e para a abertura de dados científicos. A maior parte das políticas e estratégias governamentais concentram-se na abertura dos dados governamentais em detrimento dos dados de investigação (RECODE, 2014). A expressão “Política de Dados Abertos” surge frequentemente associada a termos como estratégias, programas ou iniciativas de abertura de dados e ainda a políticas de gestão de dados. A contextualização das políticas não é clara, referindo-se apenas a Dados Abertos mas focando-se preferencialmente apenas nos dados provenientes do setor público. Em termos de dados de investigação é possível encontrar políticas de quatro intervenientes diferentes: Governo, entidade financiadora que pode ser ou não governamental, instituições de investigação e ainda as editoras de *data journals*. Em termos de dados públicos são associadas políticas governamentais e institucionais. As políticas de dados abertos devem ainda incluir uma política de gestão de dados, independentemente do seu cariz de investigação ou público.

Nos últimos anos, questões sobre a gestão, preservação e partilha de dados científicos têm sido alvo de debate público. O aumento da criação de *datasets* vastos e complexos em todo o processo científico está a criar profundos desafios para o armazenamento, acesso e interpretação de dados (EAGDA, 2014). As recentes políticas de *Open Data* pretendem disponibilizar em livre acesso os *datasets* da mesma forma que qualquer outra publicação académica, sendo para isso necessário garantir uma infraestrutura de armazenamento, acesso, pesquisa e a reutilização dessa informação (Rans, Day, Duke & Ball, 2013). As políticas são essenciais para o estabelecimento de compromissos comuns (Erway, et al., 2016). Segundo a Science International (2015:6), “effective open data in a data-intensive age can only be realised if there is systemic action at disciplinary, national and international levels”. De acordo com Tananbaum, num estudo preparado em nome da SPARC, “open data policies promote the accessibility and reuse of the raw data generated during the scientific discovery phase”. Estas políticas abrangem uma gama variada de material não-textual, como *datasets*, estatísticas, resultados de inquérito e os metadados associados a estes objetos, devendo permitir que estes sejam sujeitos a processamento por máquina e análise (Tananbaum). Existem alguns desafios técnicos, legais e éticos na criação de uma política de abertura de dados: como assegurar a heterogeneidade e interoperabilidade dos dados, a acessibilidade e pesquisabilidade, a preservação e curação, a qualidade e avaliação e também a segurança dos dados (RECODE, 2014).

A abertura de dados de investigação em todas as áreas científicas não pode ter uma abordagem *one size fits all*, pois estas produzem diferentes tipos de dados e têm

procedimentos de análise, publicação e arquivo diferenciados (Royal Society, 2012; RECODE, 2014; ERAC, 2016; Lämmerhirt, 2016). A heterogeneidade verificada entre as políticas de dados de investigação ao nível da aplicação normativa e da gestão de dados implica, de acordo com Lämmerhirt, que o cumprimento das medidas impostas pelas políticas seja irregular. O autor defende que as políticas de dados abertos devem considerar as práticas, normas, infraestruturas dentro de cada disciplina científica e também estabelecer incentivos para a publicação de dados.

A Comissão Europeia aborda a necessidade de desenvolver princípios técnicos, legais, éticos e semânticos de interoperabilidade, de assegurar a verificação e a confiança nos dados a nível local, regional e nacional, de implementar novos incentivos para a partilha e proteção de diferentes tipos de dados e, por fim, uma estrutura que possa avaliar estes princípios (Comissão Europeia: Grupo de peritos de alto nível em dados de investigação, 2010). Shadbolt (2016) ressalva que a aposta deve ser feita em políticas consistentes que determinem a forma de abertura de dados produzidos sob financiamento público, sendo essencial sublinhar os seus benefícios e partilhar as melhores práticas. Para gerar práticas e uma cultura de abertura, “we need incentives as well as mandates” (Shadbolt, 2016:2). A visão subjacente a este movimento de abertura de dados científicos é que os dados arriscam-se a ficar sem sentido se forem impedidos de ganhar valor científico, económico, social ou político (Leonelli, 2016).

Os governos devem desenvolver políticas de abertura dos dados científicos que complementem as políticas de abertura dos dados públicos, bem como apoiar a criação de *software* e recursos humanos habilitados que são críticos no sucesso de ambas as políticas (Royal Society, 2012). As políticas de Dados Abertos devem ter um carácter diferenciado das políticas de AA, devido a questões como a confidencialidade e segurança. Nesse sentido, o desenvolvimento de políticas de Dados Abertos é uma tarefa mais complexa, devendo por isso ser incorporada como uma extensão das políticas de AA (Swan, 2012). A lógica para a definição de políticas nacionais de *Open Data* encontra-se em assegurar o rigor da ciência nacional baseado na acessibilidade e reprodutibilidade dos seus resultados, em capturar o valor da abertura de dados para benefício próprio e como base da colaboração eficiente na ciência internacional (Science International, 2015, tradução livre). Embora a ciência seja um elemento central, Pasqueto, Sands e Borgman (2015) sublinham que esta lógica se centra maioritariamente no domínio económico.

Segundo a SPARC, as entidades financiadoras de investigação devem implementar políticas de Dados Abertos como forma de aumentar a disponibilidade e reutilização de dados, enquanto reduzem a duplicação de investigação, má-interpretção, fraude e manipulação de dados. As políticas devem abranger todos os dados e metadados necessários à verificação dos resultados de investigação, devendo a sua publicação ocorrer o mais rapidamente possível (Tananbaum). O tempo passado entre a criação e

a publicação de dados é uma questão importante na cultura do Open Data. Porém, a Royal Society (2012) sugere que os investigadores deveriam ter exclusividade do uso dos dados por um curto período de tempo, de maneira a poderem analisá-los e publicar os seus resultados.

A RECODE (2014) sublinha que os dados devem ser depositados num repositório que garanta a reutilização, segurança, estabilidade, estrutura de custos, relevância temática, metadados, diversos formatos de ficheiros, processamento por máquina e curadoria de dados. As políticas mais significativas das entidades financiadoras definem o AA a dados com obrigatório para a investigação financiada, com as exceções possíveis por razões éticas e legais. Requem o depósito de dados que suportam as publicações e outros dados importantes em repositórios certificados. Requerem que os investigadores descrevam os dados nos *Data Management Plan* (DMP) que são submetidos com as propostas de financiamento para serem avaliados. Para assegurar a reutilização dos dados e a habilidade de identificar e medir o cumprimento das políticas, os financiadores de investigação introduziram especificações técnicas nas suas políticas (DOI's, padrões de metadados específicos) bem como provisões para o licenciamento apropriado. Mais importante, as políticas eficientes incluem descrições claras das responsabilidades e expectativas para os intervenientes envolvidos (financiadores, investigadores, centros de dados, repositórios e instituições de investigação) (RECODE, 2014).

O sucesso das políticas dos financiadores depende da criação de um ambiente com incentivos apropriados que assegurem que os investigadores que gerem *datasets* de qualidade e os disponibilizem publicamente recebam o devido reconhecimento e recompensa, bem como o financiamento adequado para os repositórios e a monitorização dos planos de gestão de dados (Expert Advisory Group on Data Access [EAGDA], 2014). Segundo a RECODE (2014), as políticas das entidades financiadoras devem estabelecer o AA a dados de investigação por defeito, devem definir responsabilidades, conteúdos alvo, os planos de gestão de dados, tempo e local de depósito, as especificações técnicas para permitir a reutilização, o licenciamento de dados, medidas para garantir o acesso a longo termo e garantir a conformidade com a política. As instituições de investigação devem desenvolver políticas para a partilha de dados e preservação, fornecer treino e apoio aos investigadores para gerirem os seus dados de forma efetiva, criar carreiras para os gestores de dados e reconhecer os data outputs nos processos de avaliação (EAGDA, 2014). No Reino Unido, um grupo de entidades interessadas na abertura de dados assinou em Julho de 2016 a “Concordat on Open Research Data”, com o intuito de assegurar que os dados de investigação gerados e coletados por membros da comunidade de investigação britânica sejam disponibilizados de forma aberta para reutilização por outros, devendo esta abertura ocorrer de maneira consistente, “with relevant legal, ethical, disciplinary and

regulatory frameworks and norms, and with due regard to the costs involved” (Concordat on Open Research Data, 2016:1).

O desenvolvimento de políticas de abertura de dados permite que as instituições⁴³ salvaguardem o seu investimento intelectual, financeiro, humano e material e respondam à crescente pressão dos financiadores que requerem que os dados produzidos sob financiamento sejam geridos de forma apropriada e com AA (RECODE, 2014). As políticas de abertura de dados das instituições de investigação devem identificar as funções e responsabilidades de cada interveniente, tendo ainda um papel importante no desenvolvimento de *data science skills*, sendo a sua função preparar os investigadores para explorarem dados de forma efetiva, oferecer serviços de administração de dados e também a infraestrutura de computação e armazenamento necessária (RECODE, 2014; Whyte & Donnally, 2016:6). Rouder (2015) defende que as instituições devem utilizar sistemas automatizados que permitem que os dados sejam criados já abertos, eliminando erros na gestão de dados, fornecendo cópias de segurança automáticas, simplificação da partilha, aumento da disponibilidade dos dados a longo termo e proporcionando um local de partilha e armazenamento.

Segundo o RECODE (2014), os editores estão interessados na abertura de dados de investigação porque esta abertura agrega valor aos seus produtos principais (publicações tradicionais) ao melhorar a confiança da investigação publicada através da habilidade de a verificar. Os editores científicos devem estabelecer políticas para a partilha de dados e processos que permitam a contribuição de autores individuais nas publicações sejam avaliadas e que requeiram a citação⁴⁴ e o reconhecimento dos *datasets* usados no curso da publicação (EAGDA, 2014; ICSU, 2014). Os editores devem ainda requerer que se possa aceder, avaliar, usar e rastrear os dados subjacentes aos artigos, a partir dos mesmos, como condição para a publicação (Royal Society, 2012; EAGDA, 2014; ICSU, 2014). A emergência de *data journals*⁴⁵ deve ser ligada ao esforço de publicar dados de forma separada, o que permite que partes essenciais do registo científico possam ser disponibilizadas de forma inteligível para a comunidade científica (RECODE, 2014). Segundo o estudo de Sturges *et al.* (2015:2445), “a large percentage of journals have no data sharing policy and that there are inconsistencies between those that are traceable”, significando que os investigadores não sabem se devem

⁴³ Um bom exemplo encontra-se no sítio do Centro Europeu de Investigação nuclear, que apresenta três grupos de dados produzidos que podem ser reutilizados, tendo ainda definido políticas de acesso externo aos dados, reutilização e preservação para os dados provenientes de quatro experiências a decorrer com o Large Hadron Collider (Grande Colisor de Hádrons). Disponível na WWW: <http://opendata.cern.ch/about>

⁴⁴ De forma a potenciar o crédito do autor e a citação de dados, a Royal Society (2012), Altman & Crosas (2013), Carpenter (2014), Ball & Duke (2015) aconselham a criação perfis de autor, como o ORCID ou o ISNI, permitindo desta forma que os investigadores possam associar todos os seus trabalhos científicos num ponto único.

⁴⁵ Os *data journals* são plataformas de publicação de dados em AA com revisão por pares para a publicação, partilha e disseminação de dados que cobrem uma grande variedade de disciplinas (RECODE, 2014).

partilhar os dados subjacentes aos seus artigos e também não sabem como e onde depositar os dados. Contudo, o cenário parece estar a mudar com o grupo editorial Nature Springer⁴⁶ a anunciar em dezembro de 2016, o compromisso de mais de 600 *journals* com as novas políticas de partilha de dados definidas pelo grupo, encorajando ou mandatando a partilha, arquivo, citação de dados e revisão por pares (Nature Springer, 2016).

O Open Science Monitor avalia o impacto da abertura dos dados de investigação através três indicadores: número de repositórios de Dados Abertos, políticas das agências de financiamento relativas à partilha de dados e também o comportamento dos investigadores em termos de partilha de dados. O modelo de avaliação usa ferramentas como o r3data e o SHERPA/Juliet para determinar o número de repositórios e de políticas de abertura de dados, recorrendo ainda a estudos de caso para determinar o comportamento dos investigadores na partilha de dados. Foram listadas 105 políticas de financiadores na EU, sendo que 76 das instituições não possuem qualquer política de Dados Abertos, 10 encoraram o arquivo e 19 entidades financiadores requerem o arquivo de dados. Os resultados são altamente influenciados pelo enorme número de instituições de financiamento de investigação que pertencem ao Reino Unido, sendo que das 69 instituições listadas, 53 não implementaram nenhuma política de Dados Abertos. Ressalve-se também que das 19 políticas que requerem o arquivo de dados, 13 pertencem ao Reino Unido. (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2017).

O biosharing.org (2017) elaborou uma lista com as políticas de abertura de dados de investigação a nível mundial, contando com 96 referências em maio de 2017. As políticas de Dados Abertos distribuem-se por revista científica (68), entidade financiadora (22), projetos (4) e sociedade (2). Ciências da Vida e Ciências Biomédicas são os domínios com mais políticas. Não se encontra nenhuma política científica portuguesa listada.

O RECODE (2014) estabeleceu dez recomendações para a abertura de dados científicos:

- Desenvolver políticas compreensivas de AA aos dados de investigação e consistentes com as prioridades nacionais e alinhadas com os requisitos europeus;
- Assegurar o financiamento apropriado para o AA aos dados de investigação ao nível do financiamento do desenvolvimento e sustentabilidade a longo prazo das

⁴⁶ A Scientific Data, uma das revistas de AA para descrição de *datasets* científicos da divisão editorial Nature Research englobadas no grupo Nature Springer, fornece aos investigadores uma lista de repositórios recomendados para o depósito de dados consoante a sua área temática. Disponível na WWW: <https://www.nature.com/sdata/policies/repositories#social>

infraestruturas de investigação, treino de investigadores, bibliotecários e outros profissionais;

- Desenvolver políticas e iniciativas que recompensem os investigadores pelo esforço de partilharem dados de alta qualidade;
- Identificar os intervenientes-chave e as redes de colaboração para fomentar o trabalho colaborativo para um ecossistema sustentável de AA aos dados;
- Planear a sustentabilidade da curadoria e a preservação dos dados a longo termo;
- Desenvolver soluções técnicas e infraestruturais que possam arcar com a preservação de dados de alta qualidade a longo prazo: harmonização de dados, descoberta e acesso, preservação, obsolescência tecnológica, documentação e metadados, indicadores de qualidade e relevância e questões de segurança;
- Desenvolver padrões técnicos e científicos de qualidade para os dados;
- Requerer o uso de estruturas harmonizadas, abertas e licenciadas;
- Lidar com os sistemáticos desafios legais e éticos decorrentes do AA a dados;
- Apoiar a transição para os Dados Abertos através do treino e desenvolvimento de currículo.

A ERAC (2016) desenvolveu um conjunto de recomendações para apoiar o desenvolvimento de políticas nacionais de abertura de dados de investigação. As recomendações são dirigidas aos diversos intervenientes da abertura de dados, estando divididas em quatro categorias:

- Treino dos intervenientes no processo de abertura de dados: mudança cultural, treino, incentivos e sistemas de recompensa e monitorização;
- Gestão e qualidade de dados: aspetos técnicos da avaliação da qualidade de dados, identificação e citação de dados e gestão de projetos;
- Sustentabilidade e financiamento: abordagem das infraestruturas e financiamento essenciais;
- Questões legais: direitos de propriedade intelectual e questões de segurança.

A mudança cultural exigida para a abertura plena dos dados passa pela necessidade de estabelecer novas formas de reconhecimento e recompensa aos esforços de gestão, preservação e curação de dados como forma de envolver investigadores e comunidades na abertura de dados. (RECODE, 2014; International Science, 2015). A mudança cultural tem, no entanto, encontrado alguns reveses pois o caminho para a abertura de dados não é percencionado da mesma forma pelos diferentes intervenientes. Isto resulta de conflitos entre cadeias de valor, processos paralelos e desconectados, especialmente entre as práticas de cada disciplina e os requisitos de cada financiador/instituição (RECODE, 2014). Segundo o inquérito efetuado pelo EAGDA (2014), os planos de gestão de dados não são avaliados adequadamente e existe pouca monitorização da implementação da partilha de dados. Os investigadores não se sentem encorajados pelas instituições a partilhar dados e existe necessidade de

proporcionar treino aos investigadores. A falta de apoio para a partilha de dados é sintomática com a falta de habilidades na gestão de projetos, sendo que os investigadores dedicam poucos cuidados à preservação e curação de dados. Estes acontecimentos contribuem para a falta de conhecimento de metadados, limitando a utilidade dos *datasets*. Os investigadores sentem também que a partilha de dados não tem influência na progressão na carreira (EAGDA, 2014).

A OECD (2015a) considera essencial que os investigadores desenvolvam competências sobre dados que lhes permita não só partilhar dados limpos e com metadados adequados, como também entender e reutilizar dados de terceiros. A familiaridade com ferramentas e princípios de gestão de dados deve ser, segundo a Royal Society (2012), parte integrante do treino dos cientistas. As bibliotecas como parte integrante do mundo da investigação e os bibliotecários,⁴⁷ como profissionais de informação multifacetados, podem desempenhar um papel ativo no mundo dos Dados Abertos como facilitadores do processo de investigação (RECODE, 2014; Hahnel e Hook, 2016; Erway, et al., 2016; Hernández-Pérez, 2016). Para Erway, et al. (2016), o plano de gestão de dados é um ponto base para os bibliotecários fornecerem serviços de gestão de dados de investigação. As bibliotecas devem fornecer informação sobre o processo de gestão de dados incluindo planos de gestão de dados e aconselhamento sobre direitos de propriedade intelectual, treino, serviços de metadados, fornecer um catálogo de dados ou um repositório de dados, oferecer serviços de armazenamento, pesquisa e acesso permanente para apoiar o ciclo da investigação (Christensen-Dalsgaard, et al., 2012; Erway, et al., 2016; Hernández-Pérez, 2016). Christensen-Dalsgaard, et al. (2012) acrescenta que as bibliotecas devem participar no desenvolvimento da política de Dados Abertos da instituição e estabelecer colaborações com investigadores, grupos de investigação, arquivos e centros de dados para fomentar a interoperabilidade da estrutura de acesso, pesquisa e partilha de dados. Embora muitas das funções sejam idênticas, Hernández-Pérez (2016) faz a distinção entre os papéis das bibliotecas públicas e das universitárias. O autor ressalva que as bibliotecas públicas devem centrar a atenção nos Dados Abertos que sejam afetos à sua comunidade. A RDA publicou em 2015 um documento com uma visão geral sobre as práticas, recursos *online* e ferramentas que podem ajudar as bibliotecas a integrar a gestão de dados nas suas rotinas de trabalho. As 23 orientações sugeridas dividem-se em recursos de aprendizagem, disseminação de dados, planos de gestão de dados, literacia de dados, metadados, citação de dados, licenciamento e privacidades, preservação digital, repositórios de dados e comunidade de prática.

Na procura de um padrão de abertura de dados que permita potenciar a ciência e a investigação, o Data Publishing Group da FORCE11 criou os FAIR principles. Estes

⁴⁷ Veja-se neste ponto os cursos de formação sobre gestão de dados para bibliotecários oferecidos pela Universidade de Edimburgo. Disponível na WWW: <http://datalib.edina.ac.uk/mantra/libtraining.html>

princípios apresentam-se divididos em quatro facetas relacionadas, mas tecnicamente independentes (FORCE11: FAIR Data Publishing Group, 2014):



Fig. 15. Princípios FAIR Data. Fonte: Adaptado de FORCE11: FAIR Data Publishing Group, 2014. Disponível na WWW: <https://www.force11.org/group/fairgroup/fairprinciples>

A implementação destes princípios técnicos na abertura de dados poderá potenciar também a partilha e por consequente, a citação de dados. A citação de dados⁴⁸ é um assunto que tem vindo a ser debatido na comunidade científica nos últimos anos, sendo uma área em que a FORCE11 também já definiu princípios⁴⁹ que começam a ser adotados pela comunidade científica.

Relativamente aos dados públicos ou do setor público, a comunidade internacional tem vindo a aumentar o seu interesse no uso destes dados, embora nem sempre seja claro que a abertura seja uma prioridade na sua agenda política (OD4D, 2016). “The

⁴⁸ Num mundo onde o Fator de Impacto continua a ser a principal medida de avaliação do trabalho dos cientistas, a citação de dados e as métricas alternativas, denominadas de Altimetrias, começam a ganhar algum destaque na literatura científica.

⁴⁹ Os Data Citation Principles foram criados pela FORCE11 com o intuito de disponibilizar um padrão de citação de dados, cobrindo o propósito, funções e atributos da citação. O conjunto de oito princípios inclui a importância da citação de dados, a atribuição de crédito, prova de evidência, identificador único, acesso, persistência, especificidade e verificabilidade, interoperabilidade e flexibilidade. Disponível na WWW: <https://www.force11.org/group/joint-declaration-data-citation-principles-final>

culture of open data policies has to change from a culture in which opening of data is standardly considered (Zuiderwijk & Janssen, 2014:27).

Devem ser consideradas várias questões na elaboração de uma política de Dados Abertos, como a definição da sua aplicação e os objetivos, que dados irão ser publicados, quais as condições de publicação de dados, em que sítio, qual a frequência e qual a razão da publicação, devendo ainda ser calculado o impacto e os benefícios da abertura de dados (Carrara, et al., 2016a). Os objetivos das políticas de Dados Abertos podem ser sintetizados em três áreas principais: fomentar a inovação social e crescimento económico, melhorar a eficiência dos serviços do setor público e fortalecer a democracia e aumentar a transparência (Zuiderwijk, Gasco, Parycek & Janssen, 2014). Para alcançar este impacto real, os governos necessitam de implementar o fenómeno dos Dados Abertos como uma mudança cultural, assegurando que recursos humanos e materiais, infraestrutura, leis e políticas são eficientes para a implementação e extração de resultados a longo prazo (World Wide Web Foundation, 2016).

No contexto das políticas de Dados Abertos, a transparência não deve ser vista como um fim, mas sim como um meio para atingir outros objetivos como a inovação no setor público, permitindo o envolvimento democrático e fornecendo o acesso a informação, “including enhancing the findability, usability, and intelligibility of open data and providing the possibility to trace governmental processes and policy and decision making” (Zuiderwijk, et al., 2014). Dados os benefícios desejados, as barreiras à sua implementação e os baixos níveis de exploração de dados, os programas de DGA só funcionarão bem se forem desenhados mediante a apreciação da sua complexidade. O programa deve não só perceber as necessidades e as capacidades dos fornecedores de dados governamentais e privados, como também as características dos dados e a natureza dos recursos da comunidade e dos intervenientes (Dawes, Vidiasova & Parkhimovich, 2016). Num estudo comparativo de cinco políticas nacionais de Dados Abertos, Huijboom e van den Broek (2011) concluíram que faltam provas mais evidentes dos benefícios económicos, sociais e democráticos das políticas de abertura de dados governamentais. Os autores sublinham que as barreiras estão dentro das organizações governamentais, as motivações estão predominantemente fora desse espectro (Huijboom & van den Broek, 2011).

A investigação de Dawes, Vidiasova & Parkhimovich (2016) demonstra a existência de várias abordagens no desenvolvimento de programas de abertura de dados governamentais, com diversas considerações sociais e técnicas:

- Abordagem orientada a dados, ou seja, focada nas características, qualidade e disponibilidade de dados abertos;
- Abordagem orientada ao programa que foca os objetivos e as funcionalidades das estruturas e operações dos programas de DGA;

- Abordagem orientada ao uso e ao utilizador que se centra nos fatores que influenciam o uso de DGA pela comunidade;
- Abordagem de desempenho e impacto que foca uma vasta gama de considerações que podem influenciar como os programas de DGA funcionam;
- Abordagem de ecossistemas e redes que indicam os componentes de um programa de DGA e também as relações dinâmicas existentes e como estas influenciam o desempenho do programa.

O Secretário-geral das Nações Unidas através do seu grupo de aconselhamento independente efetuou recomendações para possibilitar “a data revolution in sustainable development” (The UN Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014a, tradução livre).

- Criação de normas, incentivos e regulamentos para encorajar e requerer que os dados sejam tornados publicamente disponíveis e de forma que possam ser úteis a outros utilizadores;
- Aumentar a literacia de dados para que mais pessoas usem e analisem dados;
- Criar inovações na forma em como, quando e que dados são coletados e partilhados para que estejam sempre atualizados e sejam relevantes para as pessoas e para os políticos.

Para apoiar o desenvolvimento sustentável e inclusivo, a *data revolution* deve ser guiada pelos seguintes princípios: qualidade e integridade de dados, desagregação de dados, publicação de dados de forma rápida, transparência e abertura de dados, curadoria e usabilidade de dados, proteção e privacidade, gestão e independência, recursos e capacidades tecnológicas (The UN Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014b). O repositório de dados estatísticos das Nações Unidas indexa 60 milhões de registos de *datasets* provenientes de 35 bases de dados de 20 temáticas diferenciadas. Ao inserir a palavra “Portugal” no campo de pesquisa do repositório são recuperados 2172 *datasets* distribuídos por 25 fontes diferenciadas. Estes conjuntos de dados distribuem-se por várias temáticas, sendo o setor industrial o maior fornecedor de dados com 612 *datasets* (United Nations Statistics Division, 2017).

Dawes, Vidiyasa & Parkhimovich (2016:2) defendem “the ecosystem metaphor” como uma abordagem heurística para o desenho efetivo de programas de abertura de dados governamentais, sublinhando a sua utilidade se o ecossistema for dotado de “feedback and adjustment among actors and processes”. A chave para um ecossistema de Dados Abertos funcional é a implementação de políticas que definam “the legal context, standards to facilitate data interoperability, and a stable and sustainable network for users of the data” (Donker & van Loenen, 2017).

A Sunlight Foundation publicou em 2014 as “Guidelines for open data policies”, fornecendo aconselhamento ao nível dos dados que devem ser tornados públicos e como essa abertura deve ser efetuada. Para a implementação de políticas de DGA, a Sunlight Foundation aponta como fundamental criar uma autoridade de supervisão, criar orientações e outras regulações vinculativas para a implementação da política, incorporar a perspectiva pública na implementação da política, definir prazos ambiciosos para a implementação, criar processos que assegurem a qualidade dos dados, assegurar financiamento suficiente para a implementação da política, criar ou explorar parcerias e mandar futuras revisões da política para possíveis ajustamentos (Sunlight Foundation, 2014, tradução livre).

De acordo com Dietrich (2013), uma política de dados abertos deve:

- Definir a abertura por defeito;
- Estar alinhada com os principais objetivos nacionais de desenvolvimento de políticas;
- Definir uma estratégia nacional e as suas prioridades;
- Tornar a coleta e gestão de dados numa rotina;
- Estabelecer ações de suporte ao ecossistema de dados abertos.

Segundo o Open Data GoldBook, todas as políticas de Dados Abertos devem cobrir os seguintes tópicos (Carrara, et al., 2016a, tradução livre):

- Lista das categorias de dados da organização;
- Avaliação do estatuto dos dados da organização;
- Alinhamento com os aspetos legais;
- Ações necessárias para atingir os objetivos propostos;
- Implicações técnicas;
- Implicações financeiras;
- Criação de um grupo de trabalho com responsabilidades e funções claras;
- Desenvolvimento de um plano que assegure a criação do grupo de trabalho e também a sua monitorização;
- Definição de indicadores de performance que possibilitem a medição do progresso da política relativamente aos objetivos delineados.

O Open Data Study (Hogge, 2011) aponta como razão de sucesso das estratégias de Open Data no Reino Unido e nos EUA a utilização de uma abordagem que envolve três níveis de intervenientes:

- Sociedade Civil e em particular o pequeno grupo que desenvolve aplicações e serviços a partir de dados públicos;
- Um grupo de burocratas governamentais munidos de talento e recursos que lhes permita funcionar como nível intermédio ou de ligação;
- Uma administração política recetiva à mudança e à abertura.

Em entrevista ao mesmo estudo (Hogge, 2011:10), Tim Berners-Lee sintetiza esta estratégia “It has to start at the top, it has to start in the middle, and it has to start at the bottom”, reforçando a ideia da importância destes três níveis de influência.

Através do seu Task Force Report, o e-IRG (2016) efetuou várias recomendações para a implementação de políticas de Dados Abertos: preferência pelas licenças abertas, promover a certificação dos repositórios, apresentar os dados como um bem científico, institucional e societal e por fim, ligar dados às publicações. O OD4D (2016) sublinha que os dados devem ser publicados de forma consistente, devem ser comparáveis e interoperáveis e para isso é necessário adotar padrões abertos e práticas de publicação para metadados, formatos, codificação de caracteres, identificadores, vocabulários e licenciamento.

Através do estudo efetuado às políticas de Dados Abertos de instituições holandesas, Zuiderwijk e Janssen (2014, tradução livre) efetuaram diversas recomendações que podem potenciar melhorias nas políticas de abertura de dados:

- Colaboração entre instituições usando recursos e conhecimento coletivo para apoiar a troca de conhecimentos e a superação da limitação de recursos;
- Foco no impacto, criação de valor público e não apenas na publicação de dados;
- Estímulo do uso de Dados Abertos e comunicação dos sucessos de reutilização;
- Criação de uma cultura onde a abertura de dados seja incorporada nos processos de trabalho diários.

Arzberger, *et al.* (2004:1777) elencaram uma estrutura de avaliação com cinco domínios para promover o acesso a dados: tecnológico, institucional e de gestão, financeiro e orçamental, legal e político, cultural e comportamental. Estes domínios podem ajudar na formulação de políticas de Dados Abertos, dado que são “a framework for locating and analyzing where improvements can be made”. Alguns autores destacam a pouca importância que tem sido dedicada à implementação, avaliação e alterações das políticas de Dados Abertos (Zuiderwijk, *et al.*,2014) e ao desenvolvimento de investigação que compare as mesmas políticas (Zuiderwijk & Janssen, 2014). A avaliação dos efeitos das políticas de Dados Abertos pode fornecer inferências importantes sobre as vantagens e desvantagens da abertura de dados e o valor público que estas podem criar (Zuiderwijk, *et al.*,2014). Comparar políticas de Dados Abertos em diferentes aspetos e a diferentes níveis é importante e necessário para obter “a better understanding of the common and differing elements in the policies and to identify the factors that affect the variation in policies and which influence their impact” (Zuiderwijk & Janssen, 2014:17-18). Avaliar o estado da abertura de dados pode envolver a avaliação técnica dos conjuntos de dados, das iniciativas de abertura dos dados governamentais através da medição quantitativa de resultados e impacto e julgamentos qualitativos sobre o desempenho da iniciativa de Dados Abertos (World Wide Web Foundation, 2014). Um método comum de avaliar o

efeito das políticas são as estruturas de avaliação da maturidade dos Dados Abertos, sendo que estas podem ser uma ferramenta importante na melhoria da abertura de dados e no envolvimento da sociedade. Estas estruturas são usadas como modelo para comparação de diferentes setores e/ou países, como instrumento de melhoria da qualidade e gestão dos Dados Abertos (World Wide Web Foundation, 2014; Donker & van Loenen, 2017), para entender os impactos comerciais e sociais e priorizar determinados tipos de dados (World Wide Web Foundation, 2014). A World Wide Web Foundation (2014) enfatiza que a avaliação das atividades de Dados Abertos deve ter em conta:

- Contexto em que a abertura de dados é feita: legal, organizacional, político, técnico, social e económico;
- Natureza e qualidade dos conjuntos de dados: acessibilidade, legibilidade por máquina e licenças abertas, dimensão técnica, legal, social e prática da abertura, classificação de *datasets*, avaliação da qualidade;
- Contexto de utilização: quem são os utilizadores, quais os seus propósitos e quais as atividades e modelos de negócio construídos a partir dos Dados Abertos;
- Benefícios adquiridos pela reutilização de Dados Abertos: sociais, ambientais, políticos e económicos ou comerciais.

Através de uma extensa revisão da literatura⁵⁰, Zuiderwijk e Janssen (2014) elaboraram uma estrutura de comparação para as políticas de Dados Abertos baseada em três vetores: o ambiente e contexto da política, o conteúdo da política, os indicadores de desempenho que permitem inferir o valor público alcançado. Estes vetores dividem-se em 27 indicadores de enquadramento e avaliação das políticas de Dados Abertos.

⁵⁰ As autoras mencionam publicações de Huijboom & Van der Broek (2011), Gibbs, *et al.* (2003), Eskelinen, *et al.* (2008), De Bruijn, *et al.* (2002), Dawes, *et al.* (2004), Rothenberg (2012), Public Sector Transparency Board (2012), Dawes, Pardo & Cresswell (2004), Open Government Directive (Obama, 2012), Jaeger, *et al.* (2012), Zuiderwijk, Jeffrey & Janssen (2012), Ono & Zavodny (2007), Jaeger (2007).

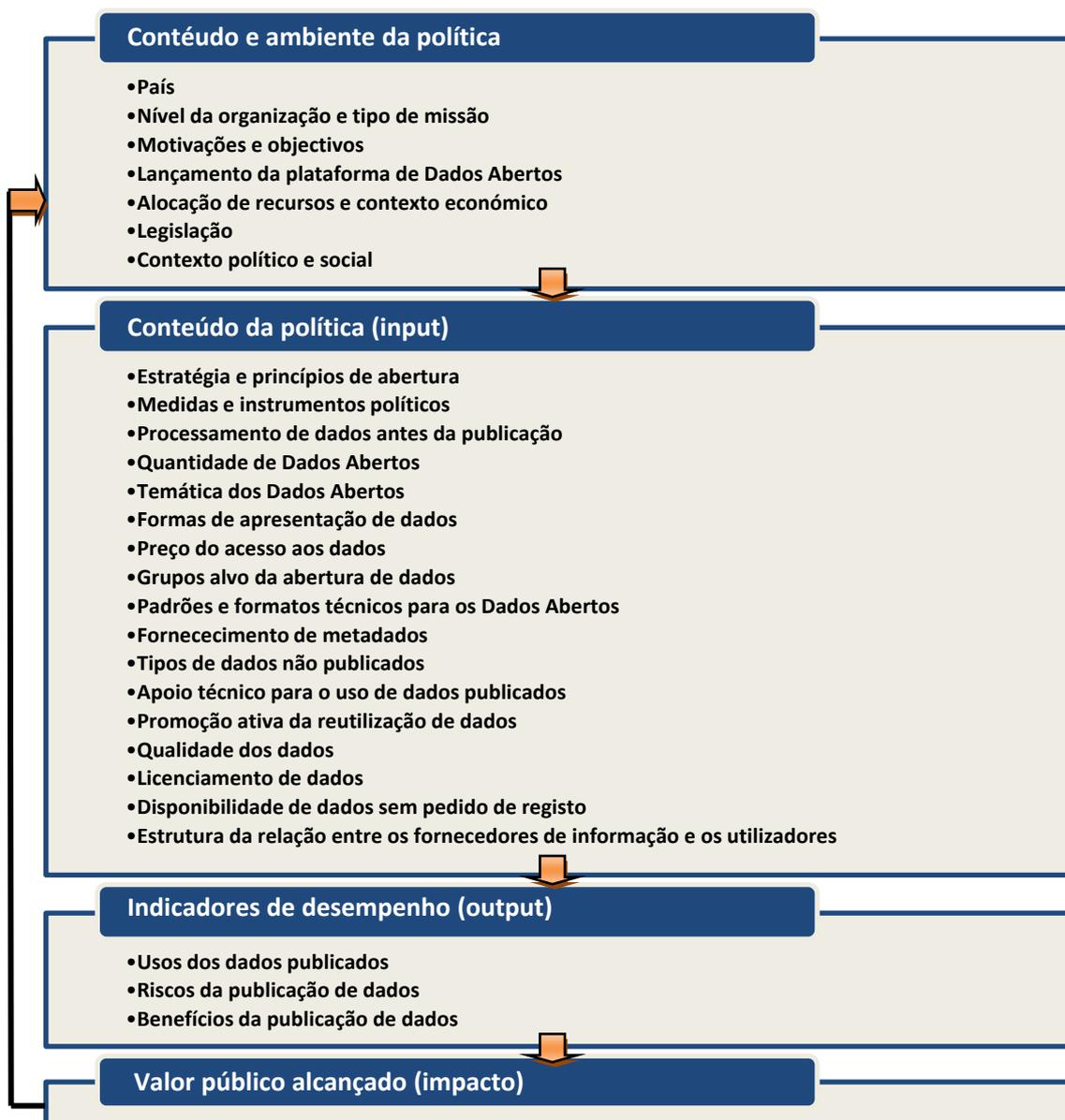


Fig. 16. Estrutura de avaliação de políticas de Dados Abertos. Fonte: Zuiderwijk e Janssen (2014). Disponível na WWW: <http://www.sciencedirect.com/science/article/pii/S0740624X13001202>

Antes de poder avaliar o sucesso do impacto dos Dados Abertos é necessário avaliar o estado atual da abertura de dados (Donker & van Loenen, 2017). De acordo com Sasse, *et al.* (2017), a avaliação dos produtos de Dados Abertos pode ser difícil. Os autores enfatizam que os responsáveis pelos portais de dados querem obter mais informação através da medição de alguns parâmetros como os *datasets* mais acedidos, que *datasets* não são atualizados ou a percentagem de *downloads*, entre outros. Estas métricas⁵¹ podem servir diversas audiências, desde organizações que publiquem dados a utilizadores que pretendam saber quais os conjuntos de dados mais usados. Ainda

⁵¹ As métricas são uma forma de avaliação que mede o progresso, desempenho, qualidade ou quantidade de um projeto, serviço, ferramenta ou produto (Sasse, *et al.*, 2017).

segundo Sasse, *et al.* (2017), as métricas devem ser integradas num contexto mais vasto para que possam conduzir à tomada de decisão e à formulação de políticas.

Entre as iniciativas de Open Data ao nível global, a disponibilização aberta de dados é normalmente feita através de catálogos e portais de dados centralizados, disponibilizando um ponto de acesso único aos utilizadores e através dos quais se pode inferir o impacto das políticas de Open Data através do regime de abertura dos *datasets* disponibilizados. O *DataCatalogs*⁵² agrega os portais de DGA espalhados pelo mundo inteiro, contabilizando 520 portais em Dezembro de 2016, sendo 197 localizados na Europa e apenas 5 em Portugal⁵³, cenário que nos parece extremamente desatualizado tendo em conta os resultados obtidos no Quadro 5.

O Open Data Barometer coleta informação sobre a abertura de dados classificando os países conforme o nível de preparação de políticas de Open Data, a sua respetiva implementação e também o impacto causado por esta abertura, oferecendo uma visão global daquilo que é o estado de abertura dos dados governamentais. A quarta edição do Open Data Barometer revela algumas conclusões preocupantes a nível da abertura dos dados governamentais. A disponibilidade de dados que cumprem todos os princípios de abertura caiu dos 10% para 7% e a legibilidade por máquina também tem tendência negativa. Apenas se registou uma melhoria no licenciamento de dados. A World Wide Web Foundation (2017) listou ainda um conjunto de problemas na disponibilização de Dados Abertos:

- Foi avaliado que nove em cada dez conjuntos de dados não estão abertos, sendo que a recomendação é que estes sejam abertos por princípio;
- Os dados governamentais são normalmente incompletos e de baixa qualidade. A World Wide Web Foundation recomenda a descentralização da abertura de dados por todas as agências e departamentos governamentais;
- A vontade política é aquilo que distingue o sucesso do insucesso da abertura de dados. “Governments must adopt the Open Data Charter to ensure open data practices are embedded beyond political mandates” (World Wide Web Foundation, 2017:17);
- Os governos não estão a publicar os dados necessários para restaurar a confiança dos cidadãos. A recomendação é que os governos devem consultar os cidadãos e intermediários para saber que dados devem ser considerados prioritários;
- Poucas iniciativas de Dados Abertos realmente promovem a igualdade e a inclusão. “Governments must invest in using open data to improve the lives of marginalised groups” (World Wide Web Foundation, 2017:21).

⁵² Disponível na WWW: <http://datacatalogs.org/>

⁵³ Os portais indexados pelo DataCatalogs são: [One.Stop.Transport](#); [iGEO](#); [Central de Dados](#); [Dados.gov.pt](#); [Despesa Pública Portugal](#)

O Open Data Maturity Model da ODI (2015:11) fornece uma estrutura para avaliar como as práticas de abertura de dados têm impacto nas organizações, nomeadamente como os dados são publicados, como são geridos e valorizados. Este modelo é baseado em cinco temas que representam uma vasta área de atividades:

- Processos de gestão de dados: identificação dos processos de gestão e publicação de dados, incluindo o controlo de qualidade, fluxos de trabalho e adoção de padrões técnicos;
- Conhecimentos e habilidades: delimitação dos passos necessários para criar “a culture of open data within an organization”, identificando a necessária partilha de conhecimento, treino e aprendizagem para entender os benefícios da abertura;
- Apoio e envolvimento do cliente: foca a necessidade das organizações de se envolver com as fontes de dados e com os reutilizadores de dados de modo a poder fornecer apoio e obter respostas para tornar a abertura de dados bem-sucedida;
- Investimento e desempenho financeiro: cobre a necessidade da organização em ter uma visão do valor dos seus dados e do orçamento necessário para suportar a sua publicação. Em termos de consumo de dados, as organizações devem entender os custos e o valor associado à reutilização de dados provenientes de outras fontes;
- Visão estratégica: sublinha a necessidade de a organização adotar uma estratégia clara sobre a partilha e reutilização de dados, e nomear uma liderança reconhecida com responsabilidade e capacidade para executar a estratégia (ODI, 2015, tradução livre).

A W3C publicou em Janeiro de 2017, como recomendação, uma lista de boas práticas para a publicação e partilha de dados denominada Data on the Web Best Practices. Estas práticas cobrem questões como o fornecimento de metadados descritivos e estruturais, licenciamento, arquivamento, enriquecimento e republicação de dados (W3C, 2017). Cada uma das práticas está classificada de acordo com os seus benefícios, definidos em oito categorias: reutilização, compreensão, conexão a outros dados, pesquisabilidade, confiança, acesso, interoperabilidade e processamento (W3C, 2017). Esta lista de boas práticas é complementada por dois vocabulários controlados, um para o uso de *datasets*⁵⁴ e outro para a qualidade dos dados⁵⁵ (W3C, 2017).

“Uma cidade inteligente é conhecedora das suas capacidades, do seu território e dos seus recursos, devendo ser capaz de transformar o seu papel de fornecedor de dados em catalisadora de processos de inovação” (Soares, Baptista e Lima, 2015). A disponibilização de dados locais permite aos cidadãos acompanharem as decisões do setor público relativas a despesas, aumentando a transparência e estimulando a democracia. Contudo, é necessário que as cidades desenvolvam uma infraestrutura correta que permita processar e integrar a gestão de dados em diversos serviços,

⁵⁴ Disponível na WWW: <https://www.w3.org/TR/vocab-duv/>

⁵⁵ Disponível na WWW: <https://www.w3.org/TR/vocab-dqv/>

criando ainda um sistema de gestão que facilite o armazenamento e publicação de dados (Carrara, et al, 2016b). A Sunlight Foundation (2013) desenvolveu o Public Policy for Public Data, um guia para ajudar as cidades a desenvolver e implementar políticas públicas de Dados Abertos, desde as razões para a abertura, passando pelos objetivos e implementação até à promoção da política.

De forma a sistematizar as iniciativas de Dados Abertos e melhorar a sua gestão, foi publicada em Espanha a UNE 178301:2015 Ciudades Inteligentes Datos Abiertos, permitindo avaliar a maturidade da publicação de Dados Abertos das cidades inteligentes através de métricas e indicadores. Esta norma define cinco grandes domínios associados aos Dados Abertos (estratégico, legal, organizacional, técnico, económico-social) fornecendo ainda vocabulários concretos para a publicação de dados (Asociación Española de Normalización y Certificación, 2015).

O World Council on City Data⁵⁶ é uma das organizações envolvidas na promoção das smart cities e da criação de padrões para os dados provenientes das cidades. Em 2014 foi publicada a norma internacional ISO 37120 com o título “Sustainable development of communities -- Indicators for city services and quality of life”, incluindo 100 indicadores que medem indicadores sociais, económicos e ambientais. Estes indicadores estão divididos em 17 categorias de dados: economia, educação, energia, ambiente, finanças, governação, fogo e emergências, respostas, saúde, recreação, segurança, abrigos, desperdícios sólidos, telecomunicações e inovação, transportes, planeamento urbano, desperdício de água e água e saneamento (Carrara, et al, 2016b). O World Council on City Data (2016) permite comparar a abertura de dados em diferentes cidades, utilizando os indicadores da ISO 3720. Em termos de cidades listadas pelo mapa desta organização, o Porto é o único representante português, destacando-se a saúde e transportes como áreas com maior abertura.

2.1 Políticas Macro

A promessa de uma ciência orientada a dados “has renewed the vigour of policy efforts to catalyse innovation” (Whyte & Donnely, 2016:1). No âmbito da abertura de dados estão a ser desenvolvidas políticas por governos, agências de financiamento e editoras científicas⁵⁷ que assegurem que os dados de investigação estejam disponíveis em repositórios públicos. A EU tem desempenhado uma dupla função na abertura de dados científicos, ao requerer a abertura de dados nos seus programas de investimento e ao coordenar a implementação de práticas de abertura de dados nos

⁵⁶ Disponível na WWW: <http://www.dataforcities.org/>

⁵⁷ O Open Access Directory possui uma lista ainda em desenvolvimento com as revistas científicas que implementaram uma política de Dados Abertos. Disponível na WWW: http://oad.simmons.edu/oadwiki/Journal_open-data_policies

estados membros (ERAC, 2016). As agências de financiamento⁵⁸ têm vindo paulatinamente a definir políticas de AA que exigem a garantia que os dados de investigação subjacentes à publicação textual são depositados em repositórios adequados de forma a permitir a verificação dos resultados de investigação e futuros uso e interpretações (UNESCO, 2015). O rápido desenvolvimento da abertura e gestão de dados científicos nos EUA e no Reino Unido são resultados de mandatos dos financiadores como o Research Council United Kingdom, National Science Foundation e da National Institutes of Health (RECODE, 2014). As instituições de financiamento reconhecem a dificuldade na mudança da cultura de partilha de dados, sendo que a tendência deverá passar por aumentar os níveis de monitorização das políticas, como a monitorização dos DMP's, auditorias institucionais, monitorização de resultados produzidos e recompensas para o apoio na gestão de dados (Whyte e Donnally, 2016).

Em 2010 a Comissão Europeia desenhou uma visão que tinha por objetivo o desenvolvimento de uma e-infraestrutura científica que potenciase o acesso, uso e reuso, bem como a qualidade dos dados, dando o impulso para a EU formular políticas para a construção da infraestrutura de investigação, dentro do enquadramento da agenda digital. Foca-se na infraestrutura necessária para a gestão de dados científicos e identifica os benefícios de acelerar o desenvolvimento de uma completa e funcional infraestrutura para os dados científicos. O relatório faz seis recomendações, destacando-se o papel governamental no desenvolvimento e uso de novas maneiras de medir o valor dos dados, a necessidade de recompensar os que contribuem para tal, treinar uma nova geração de cientistas de dados e fomentar o entendimento público (Comissão Europeia: Grupo de peritos de alto nível em dados de investigação, 2010). No mesmo documento, a Comissão Europeia salienta que existem desafios que são necessários superar na construção da e-infraestrutura, estando estes ao nível da coleta, qualidade, usabilidade, interoperabilidade, diversidade, segurança dos dados, educação e treino, publicação e acesso de dados, exploração comercial, preservação e sustentabilidade e novos paradigmas sociais.

Em 2011 a Comissão Europeia emitiu a COM (2011) 882 onde declara que a abertura de dados científicos é vital para a Europa pelo seu poder de promover negócios e oportunidades económicas, resolver problemas da sociedade e acelerar o progresso científico, sendo necessário atuar a nível local, regional, nacional e a nível europeu (Comissão Europeia, 2011).

A recomendação da Comissão Europeia sobre o acesso à informação científica e a sua preservação (2012/417/UE) pede aos Estados Membros que desenvolvam políticas e estratégias que assegurem o AA a publicações e dados de investigação provenientes

⁵⁸ A Universidade de Cambridge apresenta uma lista com as 18 entidades financiadoras a nível internacional que possuem políticas de Dados Abertos. Disponível na WWW: <http://www.data.cam.ac.uk/funders>

de financiamento público. Estas políticas devem prever os objetivos e os indicadores de avaliação, o plano de execução e o plano financeiro, garantindo que os dados resultantes de financiamento público sejam acessíveis e reutilizáveis “through digital e-infrastructures” e sejam facilmente identificados e ligados a outros conjuntos de dados. As instituições nacionais devem promover a política nacional através de mecanismos que permitam e recompensem “the sharing of research data” (EUR-Lex, 2012:41-42). A política instaurada em 2015 pela Comissão Europeia, denominada *Digital Single Market*, propôs a implementação de uma “copyright exception that would permit researchers to analyse on a large scale scientific data to which they have lawful access” (Comissão Europeia, 2016b). Esta exceção irá permitir que os investigadores europeus usem métodos de “*Text e Data Mining*” para analisar e extrair conhecimento de grandes quantidades de informação através de técnicas computacionais (Comissão Europeia, 2016a:51). Através da implementação deste *Digital Single Market*, a EU tem envidado esforços para desenvolver uma economia baseada em dados. Esta economia de dados caracteriza-se por um “ecossistema de diferentes tipos de intervenientes no mercado, como fabricantes, investigadores e fornecedores de infraestruturas, que colaboram para garantir que os dados são acessíveis e utilizáveis”. Deste modo, os intervenientes do mercado podem retirar valor desses dados, “através da criação de um leque de aplicações com um grande potencial de melhoria da vida quotidiana” (Comissão Europeia, 2017c).

Com a ambição de tornar a Europa líder mundial numa economia baseada em dados, a Comissão Europeia (2016c) pretende “disponibilizar serviços baseados na computação em nuvem e infraestruturas de dados de craveira mundial” que forneçam vantagens competitivas a investigadores, empresários e instituições públicas baseadas na utilização de grandes volumes de dados. Neste sentido, foram delineadas quatro metas a cumprir até 2020 (Comissão Europeia, 2016d):

- “A partir de 2016: Criação de uma nuvem europeia para a ciência aberta destinada aos investigadores europeus e seus colaboradores científicos”. A estratégia definida inclui a “integração e consolidação das plataformas de infraestruturas eletrónicas, da agregação das atuais nuvens científicas e infraestruturas de investigação”, e também o apoio à criação de “serviços baseados na computação em nuvem”;
- 2017: Definição de abertura, por princípio, de todos os dados científicos gerados em futuros projetos financiados pelo quadro atual de financiamento (H2020), para garantir a reutilização do volume massivo de dados produzidos pela comunidade científica;
- 2018: Lançamento de uma iniciativa para acelerar o desenvolvimento de “tecnologias quânticas, que estão na base da próxima geração de supercomputadores”;

- Até 2020: desenvolvimento e implantação, em larga escala, de uma infraestrutura europeia de computação de alto desempenho, armazenamento de dados e redes”, onde se destaca a “criação de um centro europeu de grandes volumes de dados” e a melhoria do GEANT⁵⁹, “a rede de base de investigação e inovação”.

A aposta na criação de políticas de abertura de dados, de acordo com os objetivos europeus relativamente à CA, reflete-se no atual programa de financiamento da investigação e inovação da EU, o Horizonte 2020 (H2020). Este programa criou um projeto-piloto de dados abertos com o objetivo de maximizar o acesso e reutilização de dados de investigação nas áreas científicas selecionadas⁶⁰, embora todas as outras áreas sejam convidadas a participar (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016a). Sob o artigo 29.3 do *Model Grant Agreement*, o projeto-piloto de abertura de dados requer o depósito num repositório de AA de todos os dados e metadados associados que sejam subjacentes aos resultados finais da investigação, permitindo assim o seu acesso, exploração por máquina, reprodução e disseminação sem custos para os utilizadores. Os investigadores devem fornecer informação sobre as ferramentas e instrumentos necessários para validar os resultados. Aos participantes em projetos financiados por fundos europeus é exigido a elaboração de um DMP, uma ferramenta para a gestão, publicação e reutilização de dados, permitindo descrever os conjuntos de dados gerados no processo de investigação e como e quando serão publicados. A gestão de dados deve ser efetuada segundo os princípios FAIR, tornando os dados fáceis de encontrar, acessíveis, interoperáveis e reutilizáveis (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016c).

Com a extensão do projeto-piloto a todas as áreas de investigação financiadas, a abertura de dados de investigação passou a ser a norma, por princípio. Porém, existe a opção de não participação (Comissão Europeia, 2017b), tendo sido implementado o lema “as open as possible, as closed as necessary” (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016b:8). O *opt out* ou as exceções de não participação podem ser exercidas em caso de incompatibilidade com a obrigação de proteger resultados, com razões de segurança, com regras de proteção de dados pessoais, pelo fato de o objetivo do projeto ser colocado em causa, de o projeto não gerar ou coletar dados e ainda outras razões legítimas que possam ser apresentadas (Comissão Europeia, 2017b).

⁵⁹ O GÉANT é um elemento fundamental na infraestrutura europeia, fornecendo uma rede para a criação de investigação, educação, inovação e perfeição científica. Disponível na WWW: <https://www.geant.org/About>

⁶⁰ As áreas científicas selecionadas são: Tecnologias emergentes e futuras, infraestruturas de investigação, liderança na melhoria de tecnologias industriais (tecnologias de informação e comunicação), nano e biotecnologias, segurança alimentar, sustentabilidade de recursos agrícolas, marinhos e marítimos, ambiente e eficiência de recursos, sociedades inovadoras e inclusivas, ciência de e para a sociedade e cidades inteligentes.

A abertura dos dados de investigação criados com financiamento público conta com o apoio dos projetos do OpenAIRE+ e do OpenAIRE-Connect visam estimular a abertura de dados, funcionando “as a knowledge bank and Open Access web-portal” (ERAC, 2016:67). O portal do OpenAIRE agrega em Maio de 2017 um total de 46238 conjuntos de dados, sendo que Portugal contribui com 1968 *datasets*, distribuídos por quatro entidades, embora o Instituto Politécnico de Castelo Branco seja o fornecedor de mais de 95% dos dados.

O projeto EUDAT financiado pelo H2020 visa fornecer uma infraestrutura de dados colaborativa aos investigadores europeus de qualquer disciplina científica, onde possam preservar, pesquisar aceder e processar dados, como parte da EUDAT Collaborative Data Infrastructure, uma rede de colaboração e centros de cooperação. O EUDAT fornece acesso a sete serviços diferentes relativos à gestão de dados (EUDAT[a]). Através do serviço B2FIND⁶¹ é possível pesquisar os dados armazenados através do serviço B2SHARE⁶², verificando-se a existência de 484795 conjuntos de dados (EUDAT[b]).

Vários programas de abertura de dados governamentais foram lançados em diversos países e sistemas políticos, tendo-se expandido também ao nível subnacional e local, sendo que estes programas compreendem diretivas, regras e práticas com aplicação em todos ou na maior parte dos setores do governo (Dawes, Vidiasova & Parkhimovich, 2016). Estas iniciativas são idênticas numa primeira análise, através da construção de um portal, na seleção de *datasets* semelhantes, sendo que também recorrem aos mesmos tipos de eventos, como os *hackathons*, para preparar e divulgar o lançamento de programas de DGA (Davies, 2014). Contudo, estas iniciativas diferem no seu estado de maturação, estruturas políticas e as suas prioridades públicas (Davies, 2014) e nos princípios de abertura, na quantidade e tipos de dados publicados, nos metadados fornecidos e no apoio à reutilização de dados (Zuiderwijk & Janssen, 2014). De facto, uma simples comparação dos vários índices de avaliação permitirá perceber os diferentes níveis de maturidade das políticas de Dados Abertos e a nível europeu e mundial.

Os Dados Abertos estão na agenda europeia há mais de uma década, tendo o primeiro passo sido dado em 2003 com a publicação da Diretiva 2003/98/EC sobre a reutilização da PSI (EUR-Lex, 2003a) e da Diretiva 2003/4/CE relativa ao acesso público à informação ambiental (EUR-Lex, 2003b). Esta Diretiva aplica-se à reutilização dos documentos detidos pelo setor público, porém, não obriga os estados membros a publicar dados do setor público para estimular o mercado da informação, apenas os aconselha a tornar disponíveis, nem impõe nenhuma obrigação de autorização de reutilização dos dados (Janssen, 2011). Janssen (2011) sublinha que apesar do sucesso

⁶¹ Disponível na WWW: <http://b2find.eudat.eu/dataset?q=>

⁶² Disponível na WWW: <https://b2share.eudat.eu/>

da abertura dos dados governamentais ter a contribuição da Diretiva Europeia, as iniciativas mais bem-sucedidas referem-se a princípios de transparência e prestação de contas e não à Diretiva em si.

A EU adotou ainda uma estrutura compreensiva para o acesso e reutilização de informação geoespacial, incluindo as infraestruturas eletrónicas através da INSPIRE Directive⁶³ e fomentou a disseminação do património cultural, através da Europeia, “the European digital library for cultural objects” e o acesso a informação ambiental”. (Comissão Europeia, 2013:2).

A Diretiva 2011/833/EU, sobre a reutilização de documentos publicados pela Comissão Europeia, veio autorizar, salvo nas exceções previstas, a utilização comercial e não comercial, sem custos e sem necessidade de pedido da informação produzida pela EU ou em seu nome (EUR-Lex, 2011). A Diretiva estabelece as condições de reutilização, formatos em que os documentos devem ser disponibilizados, regras de custos, transparência e não discriminação.

A Diretiva 2013/37/EU, veio reformular a autorização da reutilização da PSI. A nova diretiva requer que a informação pública seja disponibilizada “through open and machine-readable formats and together with their metadata, at the best level of precision and granularity, in a format that ensures interoperability”, incluindo agora entidades como bibliotecas, arquivos e museus na esfera da abertura de dados (EUR-Lex, 2013:3). Esta diretiva tornou obrigatória a disponibilização de informação proveniente das entidades públicas para reutilização, ao custo de reprodução, de preferência num portal e sob licenças de formatos digitais. O artigo 9.º sublinha que os Estados-Membros devem fornecer instrumentos que facilitem a pesquisa de dados, como inventários dos documentos com os respetivos metadados, disponibilizando-os, se possível, em linha, em formato legível por máquina, bem como a ligação entre o portal e o inventário. Os objetivos passam por estimular o desenvolvimento de um mercado europeu de serviços baseados na PSI, melhorar o uso e aplicação da PSI nos processos de negócio, fortalecer a competitividade no mercado interno e para resolver divergências sobre regras de reutilização entre Estados Membros.

De forma a promover a abertura dos seus próprios dados, em Junho de 2013 a EU adotou os princípios do G8 Open Data Charter, comprometendo-se a:

- Identificar e tornar disponíveis os *datasets* fundamentais ao nível da EU;
- Identificar e tornar disponíveis os *datasets* de alto valor ao nível da EU;
- Publicar os dados no EU Open Data Portal;

⁶³A INSPIRE Directive é a Diretiva 2007/2/EC do Parlamento Europeu e do Conselho de 14 Março 2007 que veio estabelecer uma Infraestrutura para a informação espacial na Comunidade Europeia. Disponível na WWW: <http://inspire.ec.europa.eu/inspire-legislation/26>

- Aplicar os princípios do G8 Open Data Charter em todos os 28 estados-membros através da revisão da Diretiva sobre a PSI e de guias de orientação;
- Apoiar atividades, consultadoria e a participação;
- Partilhar experiências na área dos Dados Abertos (Comissão Europeia, 2013).

Em 2016 entrou em vigor o novo Regulamento Geral sobre a Proteção de Dados que é relativo ao tratamento de dados pessoais e à livre circulação dos mesmos. O regulamento veio implementar o direito de pedido de esquecimento por parte do cidadão, a portabilidade de dados que permite migrar dados pessoais entre sistemas ou empresas, multas para os incumpridores, entre outras novidades. A implementação do regulamento será obrigatória até maio de 2018, sendo as legislações nacionais substituídas por uma regulamentação europeia (EUR-Lex, 2016).

A EU criou em 2012 um portal para alojar dados públicos e de investigação que provêm de organizações europeias ou de projetos financiados pela EU. O European Union Data Portal aloja dados afetos a vinte categorias contendo 10746 *datasets* em Maio de 2017, com o Eurostat a assumir a liderança destacada em termos de número de *datasets*. Os dados são classificados em domínios com o recurso a metadados controlados provenientes de vocabulários controlados como o EuroVoc⁶⁴, com a ciência (1906), emprego e condições de trabalho (1795), questões sociais (1599), a destacarem-se como os domínios com mais *datasets* publicados. Em termos de formatos de dados, os primeiros três lugares são ocupados pelos formatos ZIP (8072), HTML (7024) e text/tab-separated-values (6762). O primeiro formato estruturado, o XML, aparece apenas em quinto lugar com 595 ficheiros, seguido do primeiro formato aberto, o CSV, com apenas 360 (União Europeia, 2017). É importante tecer duas considerações sobre os formatos dos *datasets* disponibilizados. O formato ZIP é um formato de compressão de documentos e não um formato de documento em si, portanto estes 8072 ficheiros deverão ser distribuídos por outros tipos de formatos. Em segundo lugar, verifica-se uma fraca percentagem de dados a utilizarem formatos estruturados e/ou abertos que permitam a ligação com outros dados. Este portal apresenta 2663 *datasets* com ligação a Portugal, em grande parte das categorias de Ciência e do Ambiente (União Europeia, 2017).

O Public Sector Information Scoreboard é uma ferramenta “crowd sourced” utilizada para medir o estado da abertura de dados e da PSI da EU. Este painel de avaliação, através do uso de 27 indicadores, mede sete aspetos da reutilização da informação pública: implementação da Diretiva europeia referente à PSI, práticas de reuso, preço, formatos, acordos exclusivos, informação pública local e, por fim, eventos e atividades (ePSI Scoreboard, 2013).

⁶⁴ O EuroVoc é um vocabulário controlado multilingue e pluridisciplinar sobre as atividades da EU. Disponível na WWW: <http://eurovoc.europa.eu/drupal/?q=pt>

O Open Data Monitor, um projeto financiado pela Comissão Europeia que decorreu entre 2013 e 2015, teve por objetivo a extração, harmonização, medição e monitorização, análise e visualização automática de metadados dos *datasets* disponibilizados nos catálogos a nível regional, nacional e europeu (Sasse, et al., 2017). Os *datasets* “harvested from portals across Europe” (Sasse, et al., 2017:59) são estudados através de métricas quantitativas e qualitativas (Open Data Monitor, 2015). A sua plataforma permite que os utilizadores tenham uma visão dos conjuntos de dados e dos 173 catálogos de dados identificados, apresentando os resultados em termos do número de *datasets* existentes, tamanho de armazenamento, a sua disponibilidade, a posse de uma licença aberta, o uso por máquina e por fim, a exaustividade dos metadados (Open Data Monitor, 2015).

O European Data Portal, uma plataforma criada em 2015 pela EU, tem por objetivo melhorar a acessibilidade e aumentar o valor dos Dados Abertos, abrangendo toda a cadeia de valor desde a publicação de dados até à sua reutilização. Para isso coleta metadados da PSI disponível em portais de dados nos países europeus, incluindo a informação sobre a disponibilização de dados e os benefícios de reutilização, fornecendo ainda áreas dedicadas à pesquisa de *datasets*, disponibilização de dados, uso de dados, treino e biblioteca de dados. (Carrara, et al., 2016a; Carrara, Nieuwenhuis & Vollers, 2016; Sasse, et al., 2017). O European Data Portal criou o Open Data Maturity para avaliar o desenvolvimento de políticas nacionais e portais na Europa, usando a informação coletada para elaborar um índice de maturidade das políticas de Open Data baseado na preparação política e na maturidade do portal de dados (Carrara, Nieuwenhuis & Vollers, 2016). Em termos de preparação política é avaliada a presença de uma política de Dados Abertos específica, normas de licenciamento, extensão da coordenação a nível nacional, uso de dados e impacto político, social e económico. A maturidade dos portais é avaliada de acordo com a sua usabilidade, a reutilização dos dados alojados no portal e inclusão de dados de diferentes domínios. Segundo a informação apurada no relatório, a maioria dos países europeus ainda não atingiram a maturidade por completo. Em 2016 houve um aumento do número de países que completaram a sua demanda na abertura de dados, sendo agora de 56.6% entre os 31 avaliados. O número de países com uma política de Dados Abertos estabelecida passou de 23 para 28. O uso de dados chegou em 2016 aos 49.3% e o impacto da abertura de dados aos 46.3%, sendo que a maturidade dos portais estabeleceu-se nos 64.3%. Contudo, os países necessitam de elevar o reconhecimento político sobre os Dados Abertos, aumentar os processos automatizados nos seus portais para aumentar a usabilidade e reutilização dos dados e também organizar mais eventos e treinos para apoiar as iniciativas nacionais e locais (Carrara, Nieuwenhuis & Vollers, 2016). O relatório conclui que os países europeus estão a enfrentar barreiras legais, financeiras, de qualidade de dados e de políticas, sublinhando que os políticos não conhecem os benefícios da abertura de dados e que por isso, os Dados Abertos não são estabelecidos como uma prioridade política. São

deixadas ainda algumas recomendações para ajudar os países a avançar no processo de abertura de dados: implementar uma estratégia de Dados Abertos que estimula a criação de uma política, adicionar funcionalidades ao portal e melhorar a qualidade dos dados, aumentar o reconhecimento sobre os Dados Abertos, monitorizar as atividades relacionadas com o Open Data (Carrara, Nieuwenhuis & Vollers, 2016).

2.2 O caso português

O primeiro sinal de abertura científica é dado pelo artigo 50º do Decreto-Lei n.º 115/2013, de 7 de agosto, veio regulamentar a obrigatoriedade da inclusão de trabalhos académicos no RCAAP, sendo posteriormente reforçado com a entrada em vigor da Portaria n.º 285/2015, de 15 de setembro, que aprova o Regulamento Técnico de Depósito de Teses e Trabalhos de Doutoramento e de Dissertações e Trabalhos de Mestrado. Os Dados Abertos entram finalmente na política científica portuguesa em 2014. A FCT definiu e adotou uma política que visa encorajar e promover a disponibilização em “bases de dados de AA apropriadas” dos dados científicos e outros materiais de investigação criados em projetos financiados pela instituição (Fundação para a Ciência e Tecnologia [FCT], 2014). Contudo, a política da FCT “não está alinhada com os requisitos relativamente aos dados de investigação do H2020” (Comissão Executiva GT-PNCA, 2016).

Os princípios orientadores para uma política nacional de Ciência Aberta definidos em fevereiro de 2016 pelo Ministério da Ciência, Tecnologia e Ensino Superior delinearam algumas metas a curto-prazo, sendo que na componente dos Dados Abertos apenas se regista a sensibilização para a importância da partilha de dados, “como um valor essencial à ciência, especialmente no que respeita à investigação beneficiária de financiamento público” (Portugal. Ministério da Ciência, Tecnologia e Ensino Superior [MCTES], 2016:3). Foram também definidas metas a atingir até 2018, dentro das quais se destacam as que se referem a dados de investigação:

- “Cumprimento a 100% da publicação de dados resultantes de projetos com financiamento público, num repositório em acesso aberto”;
- Incluir “a verificação contínua da publicação de dados e de resultados da investigação financiada com fundos públicos” nos modelos de avaliação da CA;
- Eliminar o “duplo financiamento de publicações/dados em acesso aberto” para racionalizar os custos públicos com ciência;
- “Oferta regular de ações de formação e esclarecimento em matéria de publicação de dados e resultados de investigação em acesso aberto, propriedade intelectual e proteção de dados” (Portugal. Ministério da Ciência, Tecnologia e Ensino Superior [MCTES], 2016, 2016:4).

O acesso aos dados de investigação é finalmente incluído na política científica através da Resolução do Conselho de Ministros nº 21/2016, de 11 de abril. No âmbito da definição de uma política de Ciência Aberta, é incluída a obrigatoriedade do AA aos “dados científicos resultantes de investigação financiada por fundos públicos” e sua respetiva preservação, “por forma a permitir a sua reutilização e o acesso continuado” (Resolução do Conselho de Ministros nº 21/2016, 2016).

Em dezembro de 2016 foi criado o Grupo de Trabalho para a Política Nacional de Ciência Aberta (GT-PNCA) através do Despacho n.º 15389/2016, cujo objetivo se centra na elaboração de uma política nacional de CA e onde a abertura de dados de investigação é parte integrante. A Comissão Executiva do GT-PNCA foi dividida em quatro conjuntos de temas: acesso aberto e dados abertos, infraestruturas e preservação digital, avaliação científica e responsabilidade social científica (Despacho n.º 15389/2016 de 12 de dezembro do Ministério das Finanças e da Secretaria de Estado da Cultura e Ciência, Tecnologia e Ensino Superior, 2016).

A abertura de dados científicos está ainda numa fase precoce. O primeiro relatório da Comissão Executiva do Grupo de Trabalho para a Política Nacional de Ciência Aberta constatou que em termos de gestão e partilha de dados existem algumas iniciativas e experiências de “âmbito disciplinar, nomeadamente no contexto de projetos”, sendo expectável que o Roteiro Nacional de Infraestruturas de Investigação de Interesse Estratégico traga “progressos significativos” (Comissão Executiva GT-PNCA, 2016:7). O mesmo relatório sublinha que ao nível das instituições de investigação e de ensino superior, as iniciativas de partilha e gestão de dados são quase inexistentes.

Em termos de dados de investigação, o ecossistema é bem mais concentrado a nível disciplinar e inferior em número de e-infraestruturas, se comparado com o ecossistema de dados do setor público. Integrado no projeto RCAAP, que visa a agregação de conteúdos científicos produzidos no país, surge em 2011 o Serviço de Alojamento de Repositórios de Dados Científicos⁶⁵ (SARDC), uma plataforma de livre acesso que visa agregar os dados de investigação de instituições nacionais. Porém, com já cerca de seis anos de existência, o SARDC apresenta apenas 2007 *datasets* em Maio de 2017, que correspondem a 42 categorias de assunto. Os conjuntos disponibilizados são provenientes de apenas duas instituições: o Arquivo Português de Informação Social (APIS) com 37 e o Instituto Politécnico de Castelo Branco (IPCB) com 1970 conjuntos de dados, sendo que o SARDC apenas aloja os dados do APIS. Embora exista um espaço atribuído a mais duas instituições, o Laboratório Nacional de Engenharia Civil e Universidade Católica Portuguesa, nenhuma das entidades publicou, até agora, dados através desta plataforma. O portal nacional de dados científicos é extremamente limitado em termos de funcionalidades, permitindo apenas a buscar

⁶⁵ Disponível na WWW: <https://dados.rcaap.pt/>

por autor, data, título, assunto e tipo de documento. Não é possível perceber a legalidade da reutilização de dados pois não é referido qualquer tipo de licenciamento. A qualidade dos dados apresentados pelo SARDC é uma incógnita pela falta de indicação do uso de formatos abertos e estruturados e também de um esquema padrão de metadados. Além do SARDC, os repositórios de dados científicos existentes em Portugal são em grande parte do domínio das Ciências da Vida, como se poderá constatar no Quadro 5, com exceção do Repositório Científico do Instituto Politécnico de Castelo Branco que tem um carácter geral. Depois da criação do repositório de dados científicos e da colocação em massa de conjuntos de dados, verifica-se que o depósito de *datasets* decresceu de forma extrema.

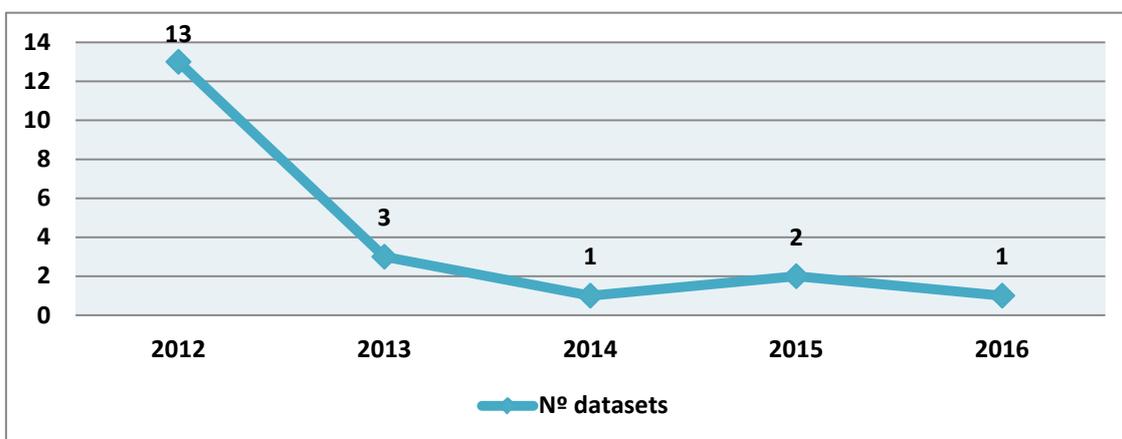


Gráfico 1. Número de *datasets* publicados no SARDC nos anos seguintes à sua criação. Fonte: Adaptado de SARDC (2017).

A tendência de abertura de dados científicos parece estar ligada à cultura disciplinar. Algumas disciplinas provenientes das ciências naturais e das ciências exatas possuem já uma estrutura uniformizada de bases de dados onde a partilha e reutilização de *datasets* se encontra mais avançada do que se verifica nas restantes disciplinas científicas. O Registry of Research Data Repositories (re3data) indexa apenas 2 repositórios de dados científicos portugueses, o INTEGRALL⁶⁶ e o KiMoSys⁶⁷, sendo que ambos pertencem ao ramo da biologia (DataCite, 2017a). O Biosharing.org (2017) indexou cinco bases de dados científicas com origem em Portugal, havendo participação portuguesa em outras duas bases de dados internacionais. Estas pertencem todas ao domínio das Ciências da Vida e da Biologia com exceção de uma, conotada com as Ciências da Terra. O DataCite⁶⁸ (2017b) criou um Digital Object Identifier (DOI) para 18632 *datasets* relacionados com o termo Portugal, sendo a grande maioria de autores portugueses e provenientes da Global Biodiversity Information Facility (GBIF), também da área da biologia. A GBIF possui um nó

⁶⁶ Disponível na WWW: <http://integrall.bio.ua.pt/>

⁶⁷ Disponível na WWW: <http://www.kimosys.org/>

⁶⁸ Os dados obtidos relativamente ao DataCite, Pangea e GitHub foram obtidos através da inscrição do termo Portugal na caixa de pesquisa de cada repositório.

português que aloja 2681051 ocorrências de biodiversidade com Portugal a fornecer de 23 conjuntos de dados que correspondem a 448136 ocorrências (Portal de Dados de Biodiversidade de Portugal, 2016). Destaca-se ainda os 2772 *datasets* sobre Portugal, muitos criados por autores portugueses, alojados no Pangea (2017) que preserva dados relacionados com as ciências da terra e do ambiente ou o GitHub (2017) que armazena 1324 registos sobre *software* relacionados com o termo Portugal.

Portugal tem um consórcio denominado ELIXIR Portugal⁶⁹ que é constituído por quatro⁷⁰ instituições de investigação que são parte da rede nacional de informação biológica - BioData.pt⁷¹. O ELIXIR Portugal (2015) “aims to provide data, tools, standards, and training in this domain and thus contribute to build an ELIXIR framework that is of added-value to all woody plant based industries”. O BioData.pt fornece acesso a duas bases de dados disciplinares, uma ontologia para descrever experiências biológicas baseadas em plantas, 4 *softwares* e 5 conjuntos de dados.

O interesse na abertura de dados de investigação tem crescido no discurso político e científico, verificando-se já a organização de uma conferência de dados de investigação e CA⁷² em 2016 e de dois fóruns de gestão de dados de investigação⁷³, o último já em 2017. Estes eventos têm a chancela direta do MCTES em coorganização com a FCT, a Universidade do Minho e FCCN, embora esta última apenas no caso da conferência. Foi lançado na Faculdade de Engenharia da Universidade do Porto o projeto TAIL com o objetivo de otimizar o processo de gestão de dados de investigação, desde a sua produção até ao depósito e partilha. Este projeto, a decorrer de 2016 até 2019, “vai construir um portfólio de exemplos de gestão de dados em diversos domínios que poderão ser usados pelos investigadores para avaliar o esforço requerido e as compensações a obter com esta atividade” (Ribeiro, et al., 2016:256-257). Os resultados serão avaliados através da utilização da plataforma Dendro “designed to help researchers use concepts from domain-specific ontologies to collaboratively describe and share *datasets* within their groups” (Universidade do Porto. Faculdade de Engenharia, 2017).

⁶⁹ Este consórcio é o nó português de uma organização intergovernamental denominada ELIXIR que pretende reunir os recursos europeus relacionados com as ciências da vida. Os recursos incluem bases de dados, *software*, material de treino, armazenamento na nuvem e supercomputadores. Disponível na WWW: <https://www.elixir-europe.org/about-us>

⁷⁰ As organizações envolvidas no BioData.pt são: Instituto Gulbenkian de Ciência, Universidade do Minho, Instituto de Biologia Experimental e Tecnológica e o Instituto de Biotecnologia e Bioengenharia.

⁷¹ O BioData.pt faz parte do Roteiro das Infraestruturas de Investigação de Portugal. Disponível na WWW: <https://www.fct.pt/apoios/equipamento/roteiro/>

⁷² A conferência “Dados de investigação e Ciência Aberta: rumo a uma estratégia nacional” apresentou-se com o objetivo de “estabelecer as linhas orientadoras da estratégia nacional para os dados de investigação”. Disponível na WWW: <https://confdados.rcaap.pt/>

⁷³ O objetivo dos fóruns sobre gestão de dados de investigação é de desenvolver competências técnicas e capitalizar saberes e práticas nos investigadores e comunidade científica. Disponível na WWW: <http://forumgdi.rcaap.pt/2forum/>

O RCAAP publicou em Abril de 2017 um documento de apoio à abertura e gestão de dados de investigação. O “Kit de Dados de Investigação” tem por objetivo prestar apoio e informação aos “gestores de repositórios da comunidade RCAAP” e fornecer ajuda semelhante a “investigadores, gestores de ciência e outros parceiros envolvidos em atividades de suporte à gestão de dados de investigação” (Carvalho, Furtado, Príncipe, 2017:5). Este documento será a base da formação que o RCAAP irá disponibilizar, fornecendo uma contextualização sobre Dados Abertos, *Big Data*, preservação digital e informação sobre: ciclos de gestão de dados, planos de gestão de dados, direitos de autor, licenciamento e privacidade de dados, políticas e diretrizes de abertura de dados de investigação, recursos de apoio e aplicação nos recursos RCAAP.

Os dados do setor público têm uma contextualização legislativa diferenciada daquela que é aplicada aos dados de investigação. O direito ao acesso à informação pública em Portugal foi estabelecido pelo artigo 268º da Constituição de 1976. Porém, apenas em 1993 o estado português legislou no sentido da abertura dos documentos administrativos, com a publicação da Lei nº 65/93, de 26 de Agosto, a Lei do Acesso aos Documentos Administrativos (LADA), posteriormente alterada pelas leis nº 8/95 de 29 de Março e 94/99, de 16 de Julho.

A evolução da conjuntura de acesso aos documentos administrativos é derivada da imposição das Diretivas Europeias, a que Portugal, como membro da EU, está obrigado. A Lei 46/2007 de 24 de Agosto virá revogar da Lei nº 65/93 e as suas alterações, “transpondo para a ordem jurídica nacional a Diretiva n.º 2003/98/CE, do Parlamento e do Conselho, de 17 de Novembro, relativa à reutilização de informações do sector público”. O artigo 24º postula a obrigação das entidades públicas em publicar documentos administrativos, de forma eletrónica, para que possam ser acedidos e reutilizados. O artigo 27º, nº 1, vem criar a Comissão de Acesso aos Documentos Administrativos (CADA) “uma entidade administrativa independente, que funciona junto da Assembleia da República e tem como fim zelar, nos termos da Lei, pelo cumprimento das disposições legais referentes ao acesso à informação administrativa” (Comissão de Acesso aos Documentos Administrativos [CADA]).

A Resolução do Conselho de Ministros n.º 91/2010 (2010:5296) veio implementar a Agenda Digital 2015. O ponto 2.3 da Resolução, no âmbito da transparência na administração pública tinha como objetivo “desenvolver um programa de administração aberta para a publicação e agregação, em formatos reutilizáveis, da informação produzida pela Administração Pública”. É neste sentido que é proposta a criação do sítio web dados.gov.pt “para publicar e agregar a informação produzida pela Administração Pública, em formatos que possam ser lidos e reutilizados por qualquer cidadão”. Esta medida teve o intuito de “facilitar o acesso dos cidadãos e empresas à informação relevante sobre os serviços públicos com o desenvolvimento de novas aplicações com valor acrescentado”. A Lei n.º 36/2011 de 21 de Junho veio

estabelecer a implementação de normas abertas nos sistemas informáticos do Estado, contribuindo assim para a adoção de formatos de dados abertos.

Em 2016, a Lei n.º 26/2016 de 22 de Agosto revogou a Lei 46/2007 de 24 de Agosto, aprovando “o regime de acesso à informação administrativa e ambiental e de reutilização dos documentos administrativos, transpondo a Diretiva 2003/4/CE, do Parlamento Europeu e do Conselho, de 28 de janeiro, e a Diretiva 2003/98/CE, do Parlamento Europeu e do Conselho, de 17 de novembro”. O artigo 3º mandata a abertura dos dados ambientais à comunidade, incluindo-os na definição de documento administrativo, devendo a sua publicação ser feita num formato aberto e legível por máquina. O artigo 10º sobre a divulgação ativa de informação prevê que sejam publicados “os documentos administrativos, dados ou listas que os inventariem que entendam disponibilizar livremente para acesso e reutilização nos termos da presente lei, sem prejuízo do regime legal de proteção de dados pessoais” Também deve ser garantida a abertura da informação cujo “conhecimento seja relevante para garantir a transparência da atividade relacionada com o seu funcionamento”, sobretudo ao nível de planos de atividades, orçamentos, relatórios de atividades e contas, balanço social e outros instrumentos de gestão similares” e também de despachos normativos internos, circulares e orientações, que comportem enquadramento estratégico da atividade administrativa” (Lei nº 26/2016 de 22 de Agosto, 2016:2780). O artigo 23º apenas prevê a reutilização dos dados para fins educativos e para fins de investigação e desenvolvimento, deixando de lado a reutilização comercial. Contudo, o dados.gov atribuiu aos dados que aloja a licença CC BY 3.0, o que não impede a reutilização comercial dos dados. É também impossível não questionar a razão de não ter sido transposta a nova Diretiva 2013/37/EU que veio revogar precisamente a Diretiva que a Lei nº 26/2016 de 22 de Agosto implementa.

A AMA (2016) defende que as políticas de Dados Abertos devem “privilegiar o interesse e pertinência dos dados para o cidadão em geral ou para setores da sociedade específicos”. Ao nível local, apenas Lisboa e Porto parecem ter definido uma política de abertura de dados. A Carta de Princípios da Política de Dados Abertos do Município de Lisboa tem por objetivo aumentar o envolvimento dos cidadãos, aumentar a transparência e o desenvolvimento económico, além da melhoria da eficiência dos serviços. A Câmara Municipal de Lisboa (CML) (2016) compromete-se a atualizar “anualmente a política de disponibilização pública de dados abertos, indicando quais os dados a libertar ao longo do ano e em que formato”. A política abrangerá todas as unidades orgânicas do município que produzam dados, sendo que a CML incentivará as Juntas de Freguesia e as empresas municipais a publicarem os seus dados. A publicação de todos os dados é feita no portal do município como ponto de acesso único, estando disponíveis sem qualquer pedido de registo. A carta de princípios sublinha que os dados devem ser atualizados, legíveis por máquina, deter uma licença aberta que permita a reutilização comercial e ainda respeitar as leis de

privacidade e segurança. A CML compromete-se a publicar inventários das coleções de dados e a promover a reutilização de dados através de iniciativas de estímulo ao desenvolvimento de aplicações e participação pública (Lisboa. Câmara Municipal [CML], 2016).

A política de Dados Abertos no Porto é enquadrada na Agenda Digital para a Europa, tendo como visão a “disponibilização e uso de dados e informação do Município do Porto, contribuindo deste modo para a competitividade do Município do Porto e da região para visitar, para residir, captar e realizar negócios no mercado nacional e internacional” (Porto. Câmara Municipal).

Utilizando a estrutura de avaliação de políticas de Dados Abertos de Zuiderwijk e Janssen (2014), foi elaborado um quadro para avaliar o estado da implementação da política de Dados Abertos em Portugal e o seu impacto. Os dados foram recolhidos com base na Lei nº 26/2016 de 22 de Agosto, na informação disponível no portal de dados nacional, através de um gestor de projeto da AMA, André Lapa, e também em alguma literatura.

Quadro 3. Visão geral sobre o contexto, conteúdo e desempenho da política de Dados Abertos em Portugal com base no esquema delineado por Zuiderwijk e Janssen (2014).

	Aspetos da política de Dados Abertos	Âmbito de aplicação
Conteúdo e ambiente da política	País	Portugal
	Nível de organização e tipo de missão	Política nacional de cariz regulamentar que é aplicada a entidades governamentais e públicas (art. 4.º)
	Motivações e objetivos	Económica e política via EU. Facilitar o acesso e reutilização da informação por parte dos cidadãos e garantir a transparência da atividade administrativa (art. 2.º)
	Lançamento da plataforma de dados	Lançamento do portal nacional em 2011
	Alocação de recursos e contexto económico	Abertura alicerçada em duas instituições públicas, a CADA, responsável pelo cumprimento da lei e a AMA que fornece a e-infraestrutura e apoio técnico. Os custos de abertura de informação do setor público são financiados pelo Plano Global Estratégico para a racionalização de custos com as TIC na Administração Pública (OECD, 2015).
	Legislação	Lei nº 26/2016 de 22 de Agosto que transpõe a Diretiva 2003/98/CE
	Contexto político e social	Cumprimento da Diretiva europeia. Disponibilização de informação útil para o cidadão

	Aspetos da política de Dados Abertos	Âmbito de aplicação
Conteúdo da política	Estratégia e princípios de abertura	Cumprimento do princípio da administração aberta que advoga que o acesso aos dados deve ser feito mediante princípios de igualdade, da proporcionalidade, da justiça, da imparcialidade. O acesso deve ser "livre e universal, bem como a acessibilidade, a interoperabilidade, a qualidade, a integridade e a autenticidade dos dados publicados e ainda a sua identificação e localização" (art. 2.º). A publicação de dados deve ser feita num formato aberto e legível por máquina (art. 3.º)
	Medidas e instrumentos políticos	Portal nacional. Não são conhecidas métricas de avaliação, nem multas ou recompensas
	Processamento de dados antes da publicação	Validação da integridade das tabelas. Não existe verificação de metadados (Lapa, 2017)
	Quantidade de dados Abertos	802 Conjuntos de dados no portal nacional. Mas a contabilidade é bastante mais elevada com a adição dos dados alojados em outros portais públicos e privados.
	Temática dos Dados Abertos	Diversa, mas com maior incidência na Economia e Finanças
	Formas de apresentação de dados	Fácil acesso e compreensão
	Preço do acesso	Gratuito
	Grupos-alvo	Developers / programadores, jornalistas de dados, investigadores / académicos, administração pública (central ou local), empresas e sociedade civil (Lapa, 2017)
Padrões e formatos técnicos	Todos os dados no dados.gov estão disponíveis em pdf, csv, xls/xlsx, e através de serviços REST (que inclui formatos como odata, JSON; ou KML e RDF, se aplicável) (Lapa, 2017)	

	Aspetos da política de Dados Abertos	Âmbito de aplicação
Conteúdo da política (cont.)	Fornecimento de metadados	Não existe um esquema padrão de metadados a ser utilizado, sendo que estes também não são verificados antes da publicação dos dados (Lapa, 2017)
	Tipos de dados não publicados	Desconhecido
	Apoio técnico para o uso de dados publicados	Sim, fornecido pela AMA mediante pedido
	Promoção ativa da reutilização de dados	A última grande atividade de promoção de reutilização, "Government as a Developer" ocorreu em 2014-2015, não se conhecendo para novas atividades previstas (Lapa, 2017)
	Qualidade dos Dados	A última medição data de 2014, onde foram encontrados alguns problemas que não foram totalmente resolvidos junto dos fornecedores (Lapa, 2017)
	Licenciamento	É aplicada a licença CC BY 3.0 a todos os <i>datasets</i>
	Disponibilidade dos dados sem registo	Todos os dados no dados.gov estão disponíveis para <i>download</i> sem necessidade de pedido ou registo
	Estrutura da relação entre fornecedores de dados e os utilizadores	Limitada

	Aspetos da política de Dados Abertos	Âmbito de aplicação
Indicadores de desempenho	Uso de dados publicados	Existem 2 aplicações elaboradas a partir de Dados Abertos disponibilizados pela AMA e Direcção-Geral das Autarquias Locais.
	Riscos da publicação de dados	Privacidade, segurança, má utilização (ainda sem casos reportados) (Lapa, 2017)
	Benefícios da publicação de dados	Fraco impacto político, económico e social segundo o Open Data Barometer. O estudo de Gomes e Soares (2014) conclui que a transparência é o único benefício real.

Valor público alcançado
Transparência governativa e disponibilidade de dados. Faltam indicadores de reutilização e estudos de impacto para uma avaliação concreta do valor alcançado.

Fonte: Elaboração própria.

Em adição às políticas de abertura de dados públicos, um dos componentes principais do ecossistema de Dados Abertos é o desenvolvimento de e-infraestruturas que possam alojar e permitir a livre reutilização de dados. O PORDATA criado em 2010 foi o primeiro portal de dados abertos em Portugal, disponibilizando dados estatísticos sobre os municípios portugueses, sobre o país e sobre a Europa. A nível público, a abertura de dados em Portugal iniciou-se com a abertura do portal de dados públicos e do portal de dados de investigação. Criado em 2011 depois da sua aprovação política, o dados.gov é uma responsabilidade da AMA (2016a:31) e “é o ponto central na estratégia nacional de Dados Abertos, constituindo-se como a porta central de acesso *online* ao repositório de dados abertos disponibilizados pelas várias instituições” públicas em Portugal. Em termos estratégicos, o gestor de projeto André Lapa declara que a AMA já assumiu a necessidade do estabelecimento de uma relação forte entre portal central e portais sectoriais, regionais e locais, sendo que a função do dados.gov deve passar por alojar dados e por apontar para dados existentes noutros portais, embora esta última ainda não tenha sido implementada (A. Lapa, comunicação pessoal, 2017). É possível retirar algumas conclusões estatísticas sobre os dados disponibilizados no dados.gov. Os 802 conjuntos de dados existentes pertencem a 26 fornecedores de dados diferenciados, embora os dados sejam maioritariamente publicados por apenas alguns dos fornecedores, com o Instituto Nacional de Estatística a fornecer metade dos dados disponibilizados, como se pode ver no gráfico seguinte:

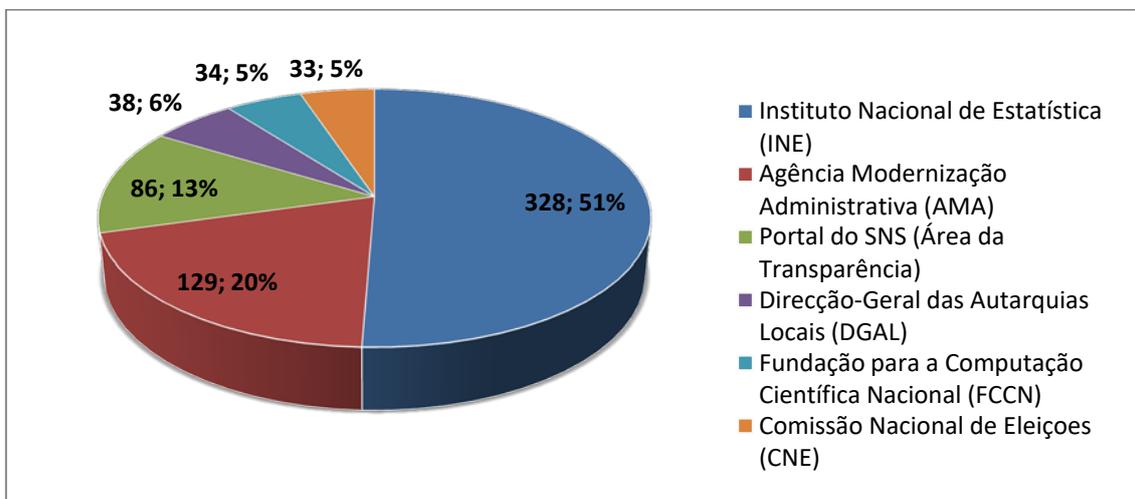


Gráfico 2. Maiores fornecedores de dados do portal nacional. Fonte: AMA (2016b)

Relativamente à tipologia dos fornecedores de dados do portal nacional, esta apresenta-se concentrada em dois tipos de entidade:

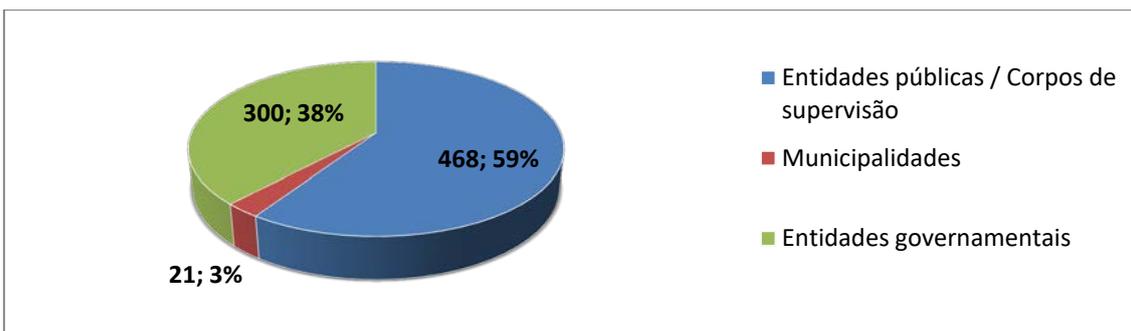


Gráfico 3. Tipologia dos fornecedores de dados do portal nacional. Fonte: AMA (2016b)

Em termos de categorias de dados disponibilizados verifica-se uma maior dispersão, embora a Economia e Finanças assumam 1/3 dos dados existentes:

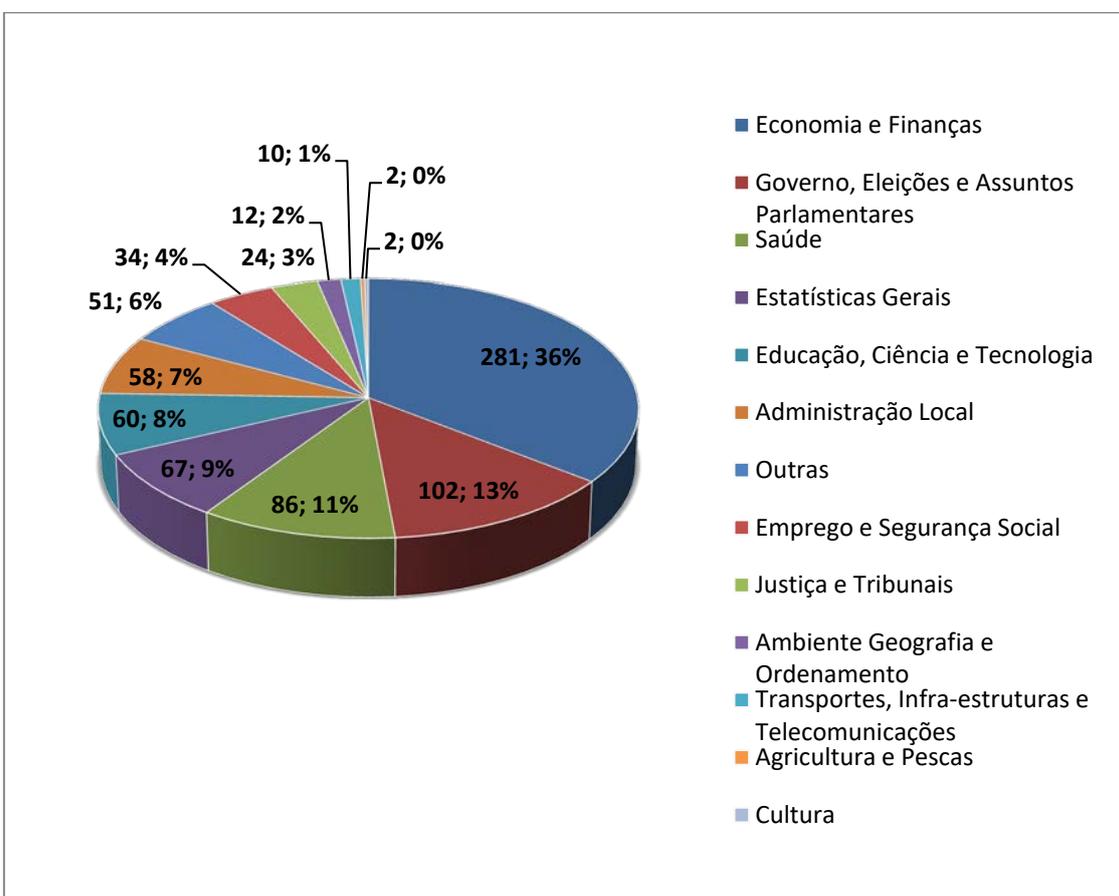


Gráfico 4. Tipologia das categorias de dados do portal nacional. Fonte: AMA (2016b)

O ecossistema de abertura de dados tem aumentado dentro e fora do espectro político. Além do portal nacional de dados públicos, um grande número de organizações públicas tem criado os seus portais de dados e/ou tem publicado os mesmos no dados.gov. Por outro lado, existem vários portais de dados com origem na sociedade civil, com o selo da Transparência Hackday Portugal⁷⁴. Organizar eventos é

⁷⁴ Transparência Hackday Portugal é um movimento que nasceu no Porto e "se define como um grupo de trabalho entusiasmado com informação pública, dados abertos e as interações possíveis entre

uma forma de promover o reconhecimento nacional sobre a abertura de dados entre organizações e cidadãos (Carrara, Nieuwenhuis & Vollers, 2016). Em Portugal têm existido diversas iniciativas a este nível, *hackathons* como o *Date with Data* organizados pela Transparência Hackday Portugal e com o apoio da Open Knowledge Portugal, com o intuito de alertar, informar e formar a comunidade para este novo paradigma na disponibilização da informação. A Open Knowledge Portugal organiza em Lisboa o *Lisbon Open Data Maker Night*⁷⁵, tendo em 2017 decorrido a terceira edição. A Porto Digital está a organizar um *hackathon*⁷⁶ sobre utilização de dados abertos no desenvolvimento de novas aplicações, que ocorrerá em Julho de 2017.

Verifica-se também que as entidades governamentais estão a apostar na criação de linhas de orientação para abertura e gestão de dados públicos. A AMA criou em 2016 um guia de Dados Abertos para apoiar a abertura dos dados públicos. O guia centra-se em quatro eixos:

- Definição de Dados Abertos e enunciação dos seus benefícios bem como as questões de privacidade;
- Potencial dos Dados Abertos a nível económico e no aumento da transparência e da participação cívica;
- Estratégia de abertura que aborda o planeamento, escolha de dados, tipos de utilizadores, dados prioritários, disponibilização de dados, estratégias de divulgação e manutenção de dados;
- Princípios que possibilitem a reutilização de dados como os formatos e as licenças abertas, além da forma de as escolher e publicar, a estrutura de dados e metadados (AMA, 2016a, tradução livre).

A AMA assume a publicação de dados como um compromisso com o cidadão, devendo a publicação ser orientada por princípios que potenciem a reutilização dos dados. A disponibilização de dados deve abranger conjuntos de dados constantes, ou sem continuidade regular, e conjuntos de dados em série, ou seja, com continuidade temporal. A relevância dos dados é justificada quando “tem potencial para ser explorado por um grande número de utilizadores ou quando o benefício potencial da sua utilização é elevado” (AMA, 2016a:20).

As duas principais cidades portuguesas, Lisboa e Porto, têm abordagens distintas relativamente às suas estratégias de abertura de dados. Lisboa possui uma parceria com a AMA e disponibiliza no seu portal⁷⁷ “*datasets* sobre a cidade, nas mais diversas áreas, de modo a potenciar a sua utilização em trabalhos de investigação ou no

tecnologia e cidadania”. O Transparência Hackday Portugal está envolvido em vários projetos de abertura de dados, entre os quais se destacam a Central de Dados e a Demo.cratika. Disponível na WWW: <http://www.transparenciahackday.org/acerca/>

⁷⁵ Disponível na WWW: <http://flipside.org/notes/open-data-maker-night-III/>

⁷⁶ Disponível na WWW: <https://hackacity.eu/porto/>

⁷⁷ Disponível na WWW: <http://dadosabertos.cm-lisboa.pt/>

desenvolvimento de aplicações informáticas” (Soares, Baptista e Lima, 2015:17). Os dados disponibilizados no portal “são produzidos no âmbito de sistemas de informação” que pertencem a diversos organismos, registando-se já diversas aplicações⁷⁸ que utilizam esses dados para criar valor para o cidadão e visitante de Lisboa na área do turismo, ambiente urbanismo (Soares, Baptista e Lima, 2015:18). Lisboa implementou também o programa Smart Open Lisboa que visa dar apoio a *start ups* na validação de “solutions using real data in a live environment and work directly with your potential customers or partners”. O apoio é dirigido a soluções na área da mobilidade, envolvimento dos cidadãos, turismo, cultura e sustentabilidade (Lisboa. Câmara Municipal, 2016). A estratégia do Porto passa pela promoção de conjuntos de Application Programming Interface (API)⁷⁹, permitindo a disponibilização de alguns serviços na web como a disponibilização de “conteúdos dos Portal de Turismo” ou do BDRT⁸⁰, entre outros, que indica pontos de interesse na cidade. A abertura de dados é centrada na “dimensão turismo e negligencia, pelo menos aparentemente, as restantes” (Soares, Baptista e Lima, 2015:18). Esta conclusão pode ser reforçada com a análise do portal de dados do Porto, onde apenas constam cinco conjuntos de dados, sendo estes pertencentes ao turismo ou relacionados, numa perspetiva de visita ou mudança como enuncia a estratégia de abertura de dados publicada pelo município.

O Porto, Lisboa, Fundão, Palmela, Penela e Águeda estão entre as cidades que integram o Open and Agile Smart Cities (OASC) (2017), uma iniciativa com o objetivo de desenvolver mercados para as cidades inteligentes baseados nas necessidades das cidades e das comunidades. O Porto é destaque nesta iniciativa pela utilização da tecnologia FIWARE⁸¹, que ajuda a melhorar o uso de dados abertos e a partilha conhecimentos e tecnologias entre cidades. A FIWARE (2015) sintetizou o impacto da abertura de dados em tempo real no Porto: “Imagine walking or driving through a city and the city itself tells you which trendy spots to check out, where to park your car in that moment, or which areas to avoid because of air quality or traffic congestion”. Na realidade, a cidade do Porto já consegue comunicar com os residentes, turistas ou empresas que utilizem “FIWARE standards and the UrbanSense platform⁸²”.

A publicação de Dados Abertos e a respetiva medição de impacto podem ser avaliadas através das várias iniciativas que visam medir o estado de abertura de dados,

⁷⁸ As aplicações referidas são: SPOT in Lisbon, Lisboa Verde, Lisboa NAV e Lisboa 360.

⁷⁹ De acordo com a Wikipédia, uma API é um conjunto de definições, protocolos e ferramentas para construir aplicações de *software*. Disponível em: https://en.wikipedia.org/wiki/Application_programming_interface

⁸⁰ Disponível na WWW: <http://bdrt.cm-porto.pt/pt>

⁸¹ A missão da FIWARE é construir um “sustainable ecosystem around public, royalty-free and implementation-driven software platform standards that will ease the development of new Smart Applications in multiple sectors”. Disponível na WWW: <https://www.fiware.org/about-us/>

⁸² A plataforma Urbansense está integrada no projeto futurecities.up da Universidade do Porto, sendo uma plataforma para a monitorização ambiental nas cidades composta por 20 unidades que coletam dados em pontos relevantes da cidade. Disponível na WWW: <http://futurecities.up.pt/site/hybrid-sensor-networking-testbed/>

algumas das quais já mencionadas, permitindo assim retirar conclusões específicas quanto ao estado dos Dados Abertos em Portugal. A análise feita por Soares, Baptista e Lima (2015) sobre a abertura de dados em quatro cidades do norte do país, Braga, Barcelos, Guimarães e Vila Nova de Famalicão, concluiu que a disponibilização de dados ao nível local é feita apenas ao nível de “contratos municipais, orçamento anual, gastos, resultados eleitorais, qualidade do ar, horários de transporte, orçamento participativo e lista de empresas”. A abertura de dados destas categorias é feita com a ajuda da “administração central e algumas instituições transversais ao território”, sendo que existem áreas onde nenhum dos municípios analisados possui Dados Abertos, como os “transportes em tempo real, instalações públicas, estatísticas criminais, segurança e higiene alimentar, acidentes de trânsito, permissões de construção, requisições de serviços e licenças comerciais” (Soares, Baptista e Lima, 2015:22). O estudo revela ainda que as cidades analisadas têm a maior taxa de abertura no Open Data Census pelo facto de serem as únicas onde a informação está completa, pois os resultados obtidos no estudo foram registados no censo.

O Portal da Transparência Municipal é uma ferramenta que permite avaliar vários indicadores da governação dos municípios portugueses com o recurso à disponibilização de dados estatísticos. Os indicadores abordam a gestão financeira e administrativa, decisões fiscais, dinâmica económica do município, serviços municipais e participação eleitoral autárquica (Portugal. Governo, 2017). Contudo, a disponibilidade de dados em cada indicador não é totalmente uniforme em termos temporais e registam-se algumas falhas de coleta ou cedência de dados.

O Open Data Census⁸³, responsabilidade da OKI, coleta informação sobre a performance de cada país e respetivas cidades em termos de abertura de dados, acessibilidade, disponibilidade, licenciamento, processamento por máquina e transparência. Existem dados de 2014 para Portugal, com um total de 178 *datasets* provenientes de 51 cidades, distribuídos por áreas como a qualidade do ar, transportes, orçamentos, contratos públicos, despesas e eleições, sendo Braga e Guimarães as cidades mais abertas. Contudo, apenas um desses conjuntos de dados possui a totalidade da abertura desejada (OKI, 2014). Para 2015 e 2016, não são apresentados resultados relativos à abertura de dados nas cidades portuguesas.

O Global Open Data Index é um índice que mede o estado de abertura dos dados governamentais ao nível global. Este índice coleta informação sobre a abertura dos dados governamentais desde 2013, ano em que Portugal se encontrava no 16º lugar, em 60 países, com uma taxa de abertura⁸⁴ calculada em 56%. Em 2014, Portugal desce

⁸³ Disponível na WWW: <http://census.okfn.org/en/latest/>

⁸⁴ O Global Open Data Index considera 9 critérios na sua avaliação para os dados serem denominados totalmente abertos: Licença aberta, dados em formatos que sejam processáveis por máquinas, acesso sem custos, disponibilidade do *dataset* completa para *download*, disponibilização de forma rápida e

para o 36º lugar em 97 países, registando uma taxa de abertura para 47%. Em 2015, nova descida do país no ranking até ao 54º lugar, num total de 122 países, com uma taxa de abertura estimada em 34%. Para o ano de 2016, as categorias das estatísticas nacionais, limites administrativos e aquisições possuem a taxa de abertura mais elevada, apenas indicando como falha o licenciamento mas que implica que nenhum dos *datasets* seja verdadeiramente aberto. A taxa de abertura calculada situa-se agora nos 37%, com Portugal a subir para 46º lugar (OKI, 2017).

O Public Sector Information Scoreboard coloca Portugal em 25º lugar num total de 31 países pontuados, sendo alguns destes resultados obtidos com base na LADA cuja última revisão data de 1999. Este painel de avaliação destaca pela positiva as práticas de reutilização, os formatos, os acordos exclusivos e os eventos e atividades. Pela negativa destaca a publicação de informação pública local a não implementação da Diretiva europeia (ePSI Scoreboard, 2013), o que não corresponde à realidade pois a Lei n.º 46/2007 de 24 de agosto, entretanto revogada pela Lei n.º 26/2016, de 22 de agosto, transpôs para a ordem jurídica nacional a Diretiva n.º 2003/98/CE.

O Open Data Barometer coloca Portugal em 2016 no 34º lugar do ranking entre 92 países, caindo três posições relativamente a 2015. Destaca-se pela positiva a prontidão da publicação de dados com 58% e pela negativa o impacto político, económico e social da abertura com apenas 16%, com uma média total de 42% de abertura. É assinalada a maior abertura de dados geográficos, censos, transportes públicos, comércio internacional, desempenho do setor da saúde, estatísticas sobre educação, crimes, ambiente e informação sobre contratos públicos. No sentido oposto, a posse de terras, informação orçamental e gastos governamentais, o registo de empresas, legislação e resultados de eleições são apontados como os dados com menos abertura (World Wide Web Foundation, 2017).

O Open Data Maturity divide os países em vários segmentos: líderes, seguidores rápidos, seguidores e iniciantes. Portugal encontra-se no grupo dos seguidores, o que indicia o desenvolvimento de uma política básica de Dados Abertos e um portal com funcionalidades avançadas. O impacto da abertura de dados é baixo, verificando-se apenas resultados positivos na transparência e eficiência política. Ao nível social é denotado um baixo impacto praticamente nulo na sustentabilidade ambiental e na inclusão de grupos marginalizados. A nível económico apenas a aplicação do modelo de custo marginal para a abertura de dados merece destaque. Em termos de preparação para a abertura de dados, Portugal desce de pontuação em relação a 2015, o que é explicado pela fase de transição para uma nova estratégia de Dados Abertos. Já a maturidade do portal nacional cresceu mais de 20% em 2016, rondando agora os 70%. Segundo os dados recolhidos pelo Open Data Maturity, o uso de Dados Abertos

imediate, disponibilidade dos dados sem conexão à internet, disponibilização dos dados em formato digital, disponibilidade pública dos dados e por fim, a confirmação da existência dos dados.

em 2016 em Portugal decaiu ligeiramente em relação ao ano anterior, rondando os 35%. O estudo indica que o país não tem uma política de Dados Abertos e que apenas vai transpor a Diretiva europeia em 2016, apesar de já o ter feito na Lei n.º 46/2007 de 24 de agosto, sendo mencionado que a estratégia a 5 anos está a ser revista (Carrara, Nieuwenhuis & Vollers, 2016). O Open Data Maturity aponta diversas barreiras em termos de publicação de dados, como a falta de apoio político para a abertura de dados valiosos e relevantes, a falta de recursos humanos, técnicos e financeiros e falta de conhecimentos relativos a padrões e boas práticas de partilha de dados públicos. O relatório afirma que “some countries have a policy framework that does not necessarily require them to add a specific licence per data set”, apontando Portugal, entre outros países, como exemplo (Carrara, Nieuwenhuis & Vollers, 2016:24). Contudo, a consulta dos avisos legais do portal dados.gov indica que todos os conjuntos de dados disponibilizados possuem uma licença aberta Creative Commons Atribuição 3.0 Portugal (CC-BY) (AMA, 2016b).

O projeto Open Data Inventory (ODIN) monitoriza o progresso das políticas de abertura dos dados públicos, medindo o seu sucesso e impacto, e fornece suporte à implementação de sistemas de dados. Fruto deste acompanhamento, o ODIN desenvolveu um ranking global com vista a avaliar a cobertura da abertura de dados e a abertura em si. De acordo com os resultados de 2016, Portugal situa-se em 15º lugar entre os 173 países cujas políticas de abertura de dados públicos foram avaliadas, com uma cobertura de 58% e abertura de 75%, perfazendo um total calculado de 67%. Este ranking foi obtido depois da avaliação da abertura de estatísticas sociais (62%), económicas (70%) e ambientais (69%) retirados do sítio web do Instituto Nacional de Estatística (INE) (Open Data Watch, 2016).

Uma iniciativa de avaliação dos programas de abertura da informação do setor público da OECD (2015b) fornece informação útil relativa ao estado dos Dados Abertos em Portugal. Segundo esta organização, os custos de abertura de informação do setor público são financiados como parte do Plano Global Estratégico para a racionalização de custos com tecnologias de informação e comunicação na Administração Pública. A legibilidade de dados por máquina é apontada como uma falha na avaliação técnica da abertura, embora André Lapa da AMA tenha confirmado que todos os dados existentes no portal dados.gov têm essa característica (A. Lapa, comunicação pessoal, 27 Março 2017). Os resultados são muitos positivos em termos de formatos abertos com a totalidade dos dados a deter esse estatuto e também na interoperabilidade, onde a OECD indica que esta característica técnica irá ser implementada como obrigatória. A nível de licenciamento, em Portugal é atribuída a Licença CC BY 3.0 à totalidade de dados alojados no portal nacional (OECD, 2015b, AMA, 2016b).

Referindo-se a dados de 2015, o portal do Open Data Monitor indica que apenas foi indexado um portal de dados em Portugal com um total de 656 datasets disponíveis,

sem, contudo, indicar qual. Em termos de resultados apurados, é indicado que Portugal apenas cumpre no capítulo da exaustividade (50%), tendo os campos da disponibilidade, uso por máquina e do licenciamento aberto com uma percentagem de 0%, o que explicará os 13% de qualidade global atribuídos pelo Open Data Monitor.

Em Maio de 2017, o catálogo do European Data Portal oferece o acesso a um total de 633680 *datasets*, provenientes de 34 países europeus. Portugal ocupa o 18º lugar com um total de 1635 conjuntos de dados disponibilizados, oriundos dos três portais indexados, o dados.gov.pt, o portal do Sistema Nacional de Informação de Ambiente (SNIAmb) e o Lisboa Aberta (European Data Portal, 2017). Verifica-se uma discrepância entre o número de *datasets* disponibilizados no catálogo do European Data Portal e o número real existente nos catálogos dos portais mencionados. O portal dados.gov.pt apresenta 802 conjuntos de dados no catálogo (AMA, 2016), enquanto o European Data Portal assinala 832 *datasets*. O SNIAmb regista um total de 1991 *datasets* (Agência Portuguesa do Ambiente, 2017) contra os 439 indexados pelo European Data Portal. Apenas os números referentes ao portal Lisboa Aberta são coincidentes, com o registo de 364 conjuntos de dados (Lisboa. Câmara Municipal, 2017).

Num estudo comparativo de seis iniciativas do norte e do sul da Europa, Gomes e Soares (2014) analisaram diversos indicadores relativos aos portais de dados nacionais, como o cumprimento dos princípios de abertura de dados governamentais, o nível de atividade do portal e os benefícios observáveis da abertura de dados. Segundo este estudo, o portal português tem nota é positiva no cumprimento dos princípios de abertura. A análise da atividade do portal é negativa, com a atividade dos visitantes, o desenvolvimento de aplicações, a atividade geradora de novo conhecimento e a utilização de recursos do portal a serem classificadas como pobres. Neste indicador apenas a atividade dos fornecedores no portal é satisfatória. Em termos de benefícios observáveis da abertura de dados governamentais, o estudo apenas regista bons níveis de transparência governativa, enquanto a participação pública, a inovação e o empreendedorismo e a eficiência governativas não revelam impactos satisfatórios.

Portugal possui ainda o Índice de Transparência Municipal (ITM) (2016) que mede o grau de transparência das Câmaras Municipais através de uma análise da informação disponibilizada nas suas páginas web. O ITM é composto por 76 indicadores divididos em sete dimensões⁸⁵, contudo, este índice é baseado no tipo de dados disponibilizados e não em que contexto de abertura estes foram publicados.

O cruzamento da informação obtida através das várias iniciativas de avaliação da abertura de dados permite inferir algumas conclusões e revelar algumas disparidades.

⁸⁵ As dimensões são: Informação sobre a Organização, Composição Social e Funcionamento do Município; Planos e Relatórios; Impostos, Taxas, Tarifas, Preços e Regulamentos; Relação com a Sociedade; Contratação Pública; Transparência Económico-Financeira; Transparência na área do Urbanismo. Disponível na WWW: <https://transparencia.pt/projetos/indice-de-transparencia-municipal/>

O Open Data Barometer e o Open Data Maturity apontam o baixo impacto económico e social da abertura de dados em Portugal. Gomes e Soares (2014) concluem que o problema em Portugal não é quantidade de dados disponibilizados mas sim a quantidade de dados que são reutilizados e como são reutilizados. O Public Sector Information Scoreboard e o Open Data Maturity afirmam que Portugal não possui uma política de Dados Abertos, o que, como vimos, não corresponde à verdade. Entre as várias iniciativas de avaliação do estado de abertura de dados, apenas três publicaram informação relativa a 2016:

Quadro 4. Comparação dos resultados de abertura de dados atribuídos pelas várias iniciativas de avaliação.

Iniciativa	Ranking	Taxa de Abertura
Global Open Data Index	46º/122 Países	37%
Open Data Barometer	34º/114 Países	42%
Open Data Inventory	15º/173 Países	67%

Fonte: Elaboração própria

Apesar da equivalência de ranking e taxa de abertura entre duas das iniciativas, é redundante inferir qualquer conclusão, dado que, estas não medem o mesmo número parâmetros, nem estes coincidem na maioria dos casos. O Open Data Inventory refere-se a dados avaliados apenas do portal do INE para calcular a sua taxa de abertura, pelo que, também deve ser relativizado.

Como pontos positivos são apontados a abertura de dados de diferentes domínios, o fato de todos os dados disponibilizados no portal estarem num formato processável por máquina (Carrara, Nieuwenhuis & Vollers, 2016), ao contrário do que afirma a OECD (2015b) e o Open Data Inventory. Apesar de a AMA não deter dados concretos do impacto da reutilização de dados existentes no dados.gov, existem pelo menos 2 aplicações, Mapa do Cidadão e Portal da Transparência Municipal, que utilizam dados do portal como fonte da informação que disponibilizam (Lapa, 2017). O Open Data Goldbook for Data Holders and Data Managers faz referência a Portugal como um exemplo de boas práticas ao nível da abertura dos dados governamentais, destacando a aplicação móvel e a plataforma *online*⁸⁶ da AMA que permite aos cidadãos aceder a um mapa que contém a localização de todos os serviços públicos (Carrara, et al., 2016a). A Infraestrutura de Dados Espaciais do Algarve disponibiliza, através do seu portal, acesso a quatro aplicações⁸⁷ que reutilizam Dados Abertos para fornecer informação de âmbito regional os cidadãos.

⁸⁶ Disponível na WWW: <https://www.mapadocidadao.pt/>

⁸⁷ As aplicações são “Sítios, Lugares e Ruas”, “Carta Regional de Equipamento”, “Reserva Ecológica Nacional” e “Planos Municipais de Ordenamento de Território”. Disponível na WWW: <http://idealg.ccdr-alg.pt/aplicacoes.aspx>

Relativamente à usabilidade do portal, o Open Data Maturity refere que não é possível contribuir para o portal, o que parece não corresponder à realidade. Citando André Lapa da AMA, organização responsável pelo portal nacional, existem fornecedores com acesso direto ao *backoffice* e que publicam os seus dados (Lapa, 2017), embora o mesmo não se verifique relativamente aos cidadãos. O Open Data Maturity avalia também a usabilidade de dados, referindo que não é possível saber a maioria dos formatos dos ficheiros existentes no portal nacional. Recorrendo à informação disponibilizada por Lapa (2017), todos os dados no dados.gov estão disponíveis em pdf, csv, xlsx, e através de serviços REST, incluindo assim formatos como odata, JSON, ou KML e RDF, se aplicável. A disponibilização dos formatos enunciados garante que todos os dados que o portal disponibiliza cumprem, pelo menos, as 3 estrelas da classificação de Tim Berners-Lee (Lapa, 2017). O Open Data Maturity afirma ainda que não existe uma licença atribuída, contudo, a AMA (2016b) enuncia que todos os seus dados estão disponíveis com uma licença CC BY 3.0. O portal dados.gov possui algumas funcionalidades como a disponibilização de uma API, que permite efetuar consultas básicas relativas ao catálogo de dados e tem capacidades de coleta automatizada para o caso dos dados disponibilizados através de *web services*, sendo que os restantes são atualizados de forma manual (Lapa, 2017).

Em Maio de 2017 contabilizam-se 60 portais de dados abertos em Portugal, distribuídos por uma variada gama de áreas. Foi elaborado um quadro com dados recolhidos através desses portais para tentar identificar as e-infraestruturas que disponibilizam Dados Abertos que cumprem os requisitos inscritos na Open Definition, nomeadamente em termos de licença, legibilidade por máquina, esquema de metadados padrão e formatos abertos. Todos os conjuntos de dados têm acesso gratuito através da web, razão pela qual esse requisito não foi incluído no quadro de avaliação. Através deste quadro foi possível retirar algumas conclusões. Existem apenas 11 portais de dados de investigação com os restantes 49 a conter dados do setor público. Os dados estatísticos são os grandes dominadores do panorama nacional, com 13 sítios web, seguidos dos dados económicos com 6. Dos 60 portais analisados apenas 33 disponibilizam dados legíveis por máquina, sendo que alguns portais não o fazem totalmente. Apenas 16 portais indicam usar uma licença aberta que permita a livre reutilização dos dados e apenas 3 fornecem a indicação do padrão de metadados usado, embora alguns dos portais pareçam usar um esquema próprio de descrição de *datasets*. Existem 17 portais que fornecem dados em formatos estruturados e 17 portais que não utilizam formatos abertos ou não indicam sequer o tipo de formato.

Quadro 5. Portais de dados abertos existentes em Portugal.

Portal	Responsabilidade	Âmbito	Total de datasets	Licença	Legib. p/ máq.	Metadados	Formatos
Repositório de dados científicos	RCAAP, FCT/FCCN	Académico	2007	IPCB – CC BY-NC APIS – CC BY 3.0	Sim	Sem indicação	IPCB – XLS APIS - PDF, XML, SAV
Repositório Científico do Instituto Politécnico de Castelo Branco	IPCB	Académico	1970	CC BY-NC	Sim	Sem indicação	XLS
Qualidade da água da Albufeira do Roxo na dinâmica dos solos e das culturas agrícolas	Assoc. Benefic. do Roxo, Inst. Politécnico de Beja, Univ. Évora	Agricultura / Recursos Hídricos	10	Sem licença específica	Sim	ISO-19139 Metadata	CSV, KML, Shapefile
Portal de dados da biodiversidade de Portugal	Instituto Superior de Agronomia	Biologia	23	CC BY 3.0, CC BY-NC 3.0	Sem indicação	Esquema próprio	Sem indicação
Biodata.pt	ELIXIR Portugal	Biologia	5	Sem indicação	Sem indicação	Sem indicação	Vários
Yeasttract	Inst. Superior Técnico - Lisboa	Ciências da Vida / Biologia	Sem indicação	Sem indicação	Sem indicação	Sem indicação	Sem indicação
Cork Oak DB	Instituto Gulbenkian de Ciência	Ciências da Vida / Biologia	Sem indicação	Sem indicação	Sem indicação	Sem indicação	HTML
Mitobreak	Universidade do Porto	Ciências da Vida / Biologia	Sem indicação	Sem indicação	Sem indicação	Sem indicação	Sem indicação
Haeckaliens	Instituto Gulbenkian de Ciência	Ciências da Vida / Biologia	Sem indicação	Sem indicação	Sem indicação	Sem indicação	Vários
Integrall	Univ. Aveiro	Ciências da Vida / Biologia	8843	Sem indicação	Sem indicação	Sem indicação	Sem indicação
Kymosys	INESC-ID Lisboa	Ciências da Vida / Biologia	42	CC BY 4.0	Sem indicação	Sem indicação	Sem indicação

Portal	Responsabilidade	Âmbito	Total de datasets	Licença	Legib. p/ máq.	Metadados	Formatos
Geodados	CM Lisboa	Informação Geográfica	257	PDDL	Sim	Esquema próprio	CSV, KLM
iGEO	APA, DGTerritório	Informação Geográfica	99	Sem indicação	Sem indicação	Esquema próprio	HTML
IDEAlg	CCDR Algarve	Informação Geográfica	Sem indicação	Sem indicação	Sem indicação	Sem indicação	Sem indicação
Mapa de Obras – Portal do Município De Vila do Bispo	CM Vila do Bispo	Autárquico	Informação geográfica	Sem indicação	Sem indicação	Sem indicação	Aplicação web
Portal de Dados Abertos do Porto	CM Porto	Autárquico	5	CC BY	Sim	Esquema próprio	SHP, XLS
Lisboa Aberta	CM Lisboa	Autárquico	365	CC0 e CC BY	Sim (não a 100%)	Esquema próprio	Diversos, a maioria são estruturados
Dados Abertos @Águeda	CM Águeda	Autárquico	47	CC0, CC BY-SA e CC BY	Sim (não a 100%)	Esquema próprio	Diversos, nem todos são estruturados
SNIAmb	APA	Ambiente	1991	Sem indicação	Sim	Esquema próprio	XML
SNIRH	APA	Ambiente / Recursos Hídricos	7210 Resultados	Sem indicação	Sim	Sem indicação	CSV, TSV
Dados.Gov	AMA	Agregador de Dados	802	CC BY 3.0	Sim	Esquema próprio	Diversos, a maioria são estruturados
Dados Abertos BNP	Biblioteca Nacional	Bibliográfico	12	CC0 1.0	Sim	UNIMARC, Dublin Core, Mods	RDF XML, RDF TURTLE e Dublin Core XML
Simplex + - Dados Abertos	AMA	Burocrático / Serviços Públicos	1	Sem licença específica, autoriza a reprodução de conteúdos	Sim	Sem indicação	JSON
Mapa do Cidadão	AMA	Burocrático / Serviços Públicos	Informação geográfica	Sem indicação	Sim	Sem indicação	Aplicação web e móvel

Portal	Responsabilidade	Âmbito	Total de datasets	Licença	Legib. p/ máq.	Metadados	Formatos
Instituto Nacional de Estatística	INE	Estatísticas Nacionais	8340	CC BY 3.0 (nos dados existentes no dados.gov)	Sim	Esquema próprio	XLS, CSV
Estatísticas Cultura	GEPAC	Cultura	Informação estatística	Sem licença específica, autoriza a reprodução não comercial de conteúdos	Não	Sem indicação	PDF
AICEP - Estatísticas	AICEP	Comércio e Investimento	Informação estatística	Sem indicação	Não	Sem indicação	PDF
BPStat	Banco de Portugal	Economia	Informação estatística	Sem indicação	Sem indicação	Esquema próprio	Sem indicação
Portugal Economy Probe	Vários	Economia	Informação estatística	Sem indicação	Sim (não a 100%)	Sem indicação	PDF, XLS
Estatísticas Economia	GEE	Economia	Informação estatística	Sem licença específica, autoriza a cópia mediante citação da fonte	Sim	Sem indicação	XLXS
Portal de Transparência Municipal	AMA e DGAL	Economia / Finanças	Informação estatística	Sem licença específica, indica que a cópia, importação ou utilização são grátis	Sim	Sem indicação	XML, CSV
Base	INCI	Economia /Compras Públicas	766326 Resultados	CC BY 3.0 (nos dados existentes no dados.gov)	Não	Esquema próprio	HTML, PDF
Estatísticas Empresariais	MJ	Estatísticas Empresariais	Informação estatística	Sem indicação	Sim	Sem indicação	PDF, XLS

Portal	Responsabilidade	Âmbito	Total de datasets	Licença	Legib. p/ máq.	Metadados	Formatos
PORDATA	Fundação Manuel dos Santos	Estatísticas Nacionais	Informação estatística	Sem indicação	Sim	Esquema próprio	HTML, PDF, XLS
PORDATA Kids	Fundação Manuel dos Santos	Estatísticas Nacionais	Informação estatística	Sem indicação	Sim	Esquema próprio	HTML, PDF, XLS
CCDR-N - Indicadores Regionais	CCDR Norte	Estatísticas Regionais	8	Sem indicação	Sim	Sem indicação	XLS
Observatório DR	CCDR Alentejo	Estatísticas Regionais	Sem indicação	Sem indicação	Sem indicação	Sem indicação	PDF, XLS
Datacentro	CCDR-Centro	Estatísticas Regionais	Informação estatística	Sem licença específica, autoriza a reprodução não comercial de conteúdos, desde que identifique a fonte	Não	Sem indicação	HTML
Estatísticas dos Açores	Governo dos Açores	Estatísticas Regionais	Informação estatística	Sem indicação	Não	Esquema próprio	HTML, PDF
Estatísticas da Madeira	Governo da Madeira	Estatísticas Regionais	Informação estatística	Sem indicação	Sem indicação	Esquema usado pelo INE	PDF, XLS
Quarenta anos de Abstenção	Transparência Hackday Portugal / Open Knowledge Portugal	Estatísticas Políticas	Informação estatística	Sem indicação	Não	Sem indicação	HTML
Central de dados	Transparência Hackday Portugal	Estatísticas Políticas e Sociais	9	PDDL	Sim	Esquema próprio	CSV, JSON
Um País como nós	UBI / U. Aveiro	Estatísticas Sociais	Informação estatística	Sem indicação	Não	Sem indicação	Aplicação web
Dados e estatísticas de cursos superiores	MEC	Educação	Informação estatística	Sem indicação	Sim (não a 100%)	Sem indicação	HTML, XLXS, ODS

Portal	Responsabilidade	Âmbito	Total de datasets	Licença	Legib. p/ máq.	Metadados	Formatos
Estatísticas Educação e Ciência	DGEEC	Educação e Ciência	Informação estatística	Sem indicação	Sim (não a 100%)	Sem indicação	PDF, XLXS, ODS
Estatísticas Justiça	DGPJ-MJ	Justiça	Informação estatística	Sem indicação	Sim (não a 100%)	Sem indicação	HTML, PDF, XLS
Sistema de informação dos certificados de óbitos	M. Saúde	Óbitos	Informação estatística	Sem indicação	Não	Sem indicação	HTML
Serviço Nacional de Saúde	SNS	Saúde	126	Sem indicação	Sim	Esquema próprio	Vários
Estatísticas Saúde	DGS	Saúde	Sem indicação	Sem indicação	Sim	Sem indicação	PDF, CSV
Dashboard da Saúde	DGS	Saúde	Informação estatística	Sem indicação	Não	Sem indicação	HTML
Relatórios Sinistralidade	ANSR	Segurança Rodoviária	Informação estatística	Sem indicação	Não	Sem indicação	PDF
Portal Opinião Pública	Fundação Francisco Manuel dos Santos	Sondagens Opinião	Informação estatística	Sem licença específica, autoriza a reprodução não comercial de conteúdos, desde que identifique a fonte	Sim	Sem indicação	HTML, PDF, XLS
Estatísticas Seg.Social	ISS	Seg. Social	Informação estatística	Sem indicação	Não	Sem indicação	HTML, PDF
O Preço Certo em Votos!	Transparência Hackday Portugal / Open Knowledge Portugal	Estatísticas Políticas	Informação estatística	General Public License v2.0	Sem indicação	Sem indicação	HTML
Arquivo Português de Informação Social	APIS	Informação social	37	CC BY 3.0	Sim	Padrão do DDI	PDF, XML, SAV

Portal	Responsabilidade	Âmbito	Total de datasets	Licença	Legib. p/ máq.	Metadados	Formatos
Mila	Transparência Hackday Portugal	Aplicação	78 Resultados	Sem indicação	Sim	Sem indicação	Aplicação web
GeoDevolutas	Transparência Hackday Portugal	Aplicação	250 Resultados	Sem indicação	Sim	Sem indicação	Aplicação web (também disponível em CSV)
Open Food Facts	Open Food Facts - Portugal	Aplicação	Informação estatística e descritiva	CC BY 4.0	Sim	Sem indicação	Aplicação web e móvel
Invasão dos hamburques	Transparência Hackday Portugal	Aplicação	Informação geográfica	Sem indicação	Sim	Sem indicação	Aplicação web (também disponível em PDF, CSV, ODS, XLSX, TSV)

Conclusão

O conhecimento é inseparável do progresso científico, do processo de decisão, da inovação e do bem-estar social. Apesar de não ser objetivo principal deste estudo, não deixámos de constatar que o ecossistema português de AA parece estar a amadurecer. Além do RCAAP, existem largas dezenas de repositórios pertencentes a universidades e institutos de investigação, alguns incluídos no ranking *Top Portals*. O ROARMAP lista 21 políticas de AA por via verde em instituições do ensino superior, o DOAJ (2017) indexa 78 revistas de AA com 91% a publicarem sob uma licença CC que permite a sua livre reutilização. O SHERPA/RoMEO listou 127 políticas editoriais de AA em Portugal, sendo que mais de metade autoriza a publicação em AA do *posprint* ou da versão final do editor. Nota ainda para a existência de uma política de AA da FCT por via do artigo 29.2 do acordo de financiamento, que mandata a abertura de todos as publicações científicas produzidos sob financiamento do programa H2020. Contudo, os poucos resultados obtidos através da literatura explorada não permitem inferir a dimensão cultural da implementação do AA em Portugal.

A plena implementação do conceito de CA em Portugal parece ainda longe da realidade. O estabelecimento de uma política de CA apenas começou a ser desenhado em fevereiro de 2016 com a publicação dos princípios orientadores para a “Ciência Aberta, Conhecimento para todos” pelo MCTES. Em Maio, a Resolução do Conselho de Ministros nº 21/2016 definiu os princípios da implementação da política de CA em Portugal, estruturada em torno de quatro eixos: AA e Dados Abertos; Infraestruturas e preservação digital; Avaliação científica; Responsabilidade social científica (Comissão Executiva GT-PNCA, 2016).

Hoje aplicamos o termo Dados Abertos referindo-nos a dados de investigação e a dados do setor público. A origem da abertura destes dois tipos de dados tem contextos muito diferenciados e independentes, embora os objetivos possam ser semelhantes. Os dados sempre foram essenciais em muitas áreas da investigação e nos últimos anos tornaram-se destaque em mais áreas científicas, crescendo substancialmente em escala e complexidade. A nível mundial verifica-se uma atenção crescente para o fenómeno da abertura de dados de investigação, sobretudo quando se debate a resolução de problemas à escala global. Existem também iniciativas e organizações internacionais envolvidas na promoção da abertura de dados que procuram fomentar o desenvolvimento de sistemas e procedimentos a nível internacional que assegurem o acesso, sustentabilidade e interoperabilidade. Muitas destas organizações já desenvolveram princípios de abertura e protocolos de dados próprios, formaram grupos de trabalho para a definição de padrões de gestão, publicação, citação, interoperabilidade e partilha de dados, ou criaram cursos de treino em gestão de dados de investigação para oferecer aos seus investigadores. Analisando em detalhe os

princípios propostos para a abertura dos dados de investigação verifica-se que, além da grande variedade de princípios propostos, a abertura legal e técnica dos dados só é acautelada em metade das recomendações. Destacam-se o AA aos dados e a reutilização de dados em número de menções. Em contrapartida, a citação de dados, a interoperabilidade, o acesso sem discriminação, as licenças e formatos abertos, a modificação e legibilidade por máquina são mencionados apenas em três recomendações, tendo a seu favor o facto de terem sido incluídas nas recomendações mais recentes. Princípios relevantes como a publicação imediata, a qualidade e a integridade dos dados surgem apenas duas vezes listados. Verifica-se também o lançamento de revistas científicas específicas para a ciência orientada a dados possibilitando a publicação de dados e dos artigos científicos que os utilizam após o serviço de revisão por pares.

Apesar do progresso, a literatura examinada aponta no sentido de que a abertura a dados de investigação não ser ainda um comportamento comum, sendo necessário uma mudança cultural (Royal Society, 2012; Pontika et al., 2015; Science International, 2015). Verificam-se três fatores que parecem estar a atrasar esta mudança no comportamento dos investigadores:

- a) A OKI (2012b) coloca os dados de investigação ao mesmo nível de importância que outros tipos de dados, como os dados culturais, financeiros, estatísticos, ambientais e ainda sobre o clima e transportes, sem lhes atribuir qualquer destaque ou importância especial.
- b) A avaliação e revisão por pares dos dados não têm a mesma qualidade quando comparada com as publicações tradicionais, dado a publicação de dados não ser ainda uma prática comum (ERAC, 2016).
- c) Hahnel e Hook (2016) apontam a falha na atribuição de incentivos e recompensas tangíveis aos investigadores pelas suas contribuições. A publicação de dados de forma aberta envolve tarefas que consomem muito tempo aos investigadores como a mudança do formato dos dados, manter a informação atualizada, alinhar os dados com as licenças existentes e com os critérios de publicação de informação sensível (Reitano, 2013).

O debate sobre a abertura dos dados governamentais tem mais de uma década. Diversas iniciativas e organizações internacionais têm promovido princípios de abertura de dados do setor público, como a OKI (2012a), que identificou as componentes chave da abertura de dados públicos: disponibilidade e acesso reutilização e redistribuição, participação universal. Foram analisados princípios de abertura de dados públicos de nove organizações e tal como em relação aos dados de investigação verifica-se uma grande variedade de princípios propostos. Embora se registre uma melhoria significativa na abertura legal e técnica relativamente aos dados de investigação, também os princípios propostos para os dados públicos não abarcam

a totalidade desta dupla abertura. O acesso, a publicação imediata, a reutilização e a disponibilização de dados de forma completa, sem discriminação e de forma a serem processados por máquina são os princípios que reúnem mais consenso entre as recomendações analisadas. As licenças abertas, a comparabilidade de dados, os formatos abertos, a disponibilidade, a abertura por defeito de dados, a interoperabilidade, publicação de metadados e a gestão de dados após a sua publicação são mencionados em menos de metade das recomendações analisadas, apesar de ser indiscutível a sua importância no processo de abertura de dados.

A abertura dos dados governamentais enfrenta restrições complexas a nível institucional, da complexidade da tarefa, uso e participação, legislação, qualidade da informação e questões técnicas (Janssen, Charalabidis e Zuiderwijk, 2012). Os governos devem modificar as práticas e procedimentos internos de gestão de informação, pois a reutilização da informação está dependente da forma como é gerida e publicada (Reitano, 2013). A OECD (2015a) declara que estes dados continuam muitas vezes inacessíveis, quer em termos de acesso, quer em termos de custo. São apontados alguns entraves à sua publicação derivados de questões culturais (Carrara, *et al.*, 2016a), restrições técnicas e financeiras (Ubaldi, 2013; Carrara, *et al.*, 2016a), sendo necessário enfrentar desafios relacionados com políticas, tecnologias, finanças, organização e cultura para criar um ecossistema e construir um modelo de negócios sustentável para as iniciativas de DGA (Ubaldi, 2013, tradução livre). Por outro lado, devem ser fornecidos aos cidadãos os conhecimentos e habilidades necessárias para interpretar a vasta gama de conjuntos de dados e gerar valor a partir desses dados (Janssen, 2012; Koski, 2015).

O impacto da abertura dos dados governamentais pode ser medido em três dimensões: económica, política e social (Granicas, 2013; Koski, 2015). Koski (2015:3) defende ser necessário desenvolver um modelo de avaliação e monitorização seguido da coleta sistemática de dados para medir com exatidão o impacto dos Dados Abertos.

A abertura de dados sem restrições de acesso deve ser garantida por uma política de Dados Abertos, pois “Open Data is political” (World Wide Web Foundation, 2016:38). O fenómeno dos Dados Abertos é também técnico e social, dada a interação entre o meio local, político, social, legal e técnico e as tecnologias que produzem resultados (Davies, 2014; Pasqueto, Sands & Borgman, 2015). A literatura destaca a existência de políticas diferenciadas para a abertura de dados governamentais e para a abertura de dados científicos, com maior enfoque político na abertura dos dados governamentais (RECODE, 2014). Em termos de dados de investigação é possível encontrar políticas de quatro intervenientes diferentes: Governo, entidade financiadora que pode ser ou não governamental, instituições de investigação e ainda as editoras de *data journals*. Ao nível dos dados públicos são associadas políticas

governamentais e institucionais. As políticas de dados abertos devem ainda incluir uma política de gestão de dados, independentemente do cariz dos dados.

As recentes políticas de *Open Data* pretendem disponibilizar em livre acesso os *datasets* da mesma forma que qualquer outra publicação académica, sendo para isso necessário garantir uma infraestrutura de armazenamento, acesso, pesquisa e a reutilização dessa informação (Rans, Day, Duke & Ball, 2013). Segundo a Science International (2015:6), “effective open data in a data-intensive age can only be realised if there is systemic action at disciplinary, national and international levels”.

Exige-se uma mudança cultural para atingir a abertura plena dos dados de investigação, sendo necessário estabelecer novas formas de reconhecimento e recompensa aos esforços de gestão, preservação e curadoria de dados como forma de envolver investigadores e comunidades na abertura de dados. (RECODE, 2014; Internacional Science, 2015). É essencial que os investigadores desenvolvam competências que lhes permita não só partilhar dados com metadados adequados, como também entender e reutilizar dados de terceiros (OECD, 2015a). A familiaridade com ferramentas e princípios de gestão de dados deve ser, segundo a Royal Society (2012), parte integrante do treino dos cientistas. As bibliotecas e os bibliotecários como profissionais de informação multifacetados podem desempenhar um papel ativo no mundo dos Dados Abertos como facilitadores do processo de investigação (RECODE, 2014; Hahnel e Hook, 2016; Erway, et al., 2016; Hernández-Pérez, 2016).

A implementação de uma política de abertura de dados públicos deve ter em conta a definição da sua aplicação e os objetivos, que dados irão ser publicados, quais as condições de publicação de dados, em que sítio, qual a frequência e qual a razão da publicação, devendo ainda ser calculado o impacto e os benefícios da abertura de dados (Carrara, et al., 2016a). “The culture of open data policies has to change from a culture in which opening of data is standardly considered” (Zuiderwijk & Janssen, 2014:27). Os governos necessitam de implementar o fenómeno dos Dados Abertos como uma mudança cultural, assegurando que recursos humanos e materiais, infraestrutura, leis e políticas são eficientes para a implementação e extração de resultados a longo prazo (World Wide Web Foundation, 2016). A literatura indica algumas estruturas de avaliação da abertura de dados públicos, tendo o nosso enfoque recaído no modelo criado Zuiderwijk e Janssen (2014), que aborda o contexto, conteúdo e indicadores de desempenho da política de dados abertos para calcular o valor alcançado pela abertura de dados. Destaque ainda para o Open Data Barometer que revela algumas conclusões preocupantes a nível da abertura dos dados governamentais. A disponibilidade de dados que cumprem todos os princípios de abertura caiu dos 10% para 7% em 2016 e a legibilidade por máquina também tem tendência negativa. Apenas se registou uma melhoria no licenciamento de dados.

Ao nível macro, a EU tem estado ativa na recomendação e implementação de políticas de abertura de dados de investigação e governamentais. Em 2015, no âmbito da criação do Mercado Digital Único, propôs a implementação de uma “copyright exception that would permit researchers to analyse on a large scale scientific data to which they have lawful access” (Comissão Europeia, 2016b). Com a ambição de tornar a Europa líder mundial numa economia baseada em dados, a Comissão Europeia (2016c) pretende “disponibilizar serviços baseados na computação em nuvem e infraestruturas de dados de craveira mundial” que forneçam vantagens competitivas a investigadores, empresários e instituições públicas baseadas na utilização de grandes volumes de dados.

A aposta na criação de políticas de abertura de dados, de acordo com os objetivos europeus relativamente à CA, reflete-se no atual programa de financiamento da investigação e inovação da EU, o Horizonte 2020 (H2020). Foi criado um projeto-piloto de dados abertos com o objetivo de maximizar o acesso e reutilização de dados de investigação nas áreas científicas selecionadas, tendo sido expandido em 2017 a todas as áreas sob financiamento do H2020. O artigo 29.3 do *Model Grant Agreement*, requer o depósito num repositório de AA de todos os dados e metadados associados, permitindo assim o seu acesso, exploração por máquina, reprodução e disseminação sem custos para os utilizadores. Os investigadores devem também fornecer informação sobre as ferramentas e instrumentos necessários para validar os resultados e elaborar um Data Management Plan para descrever os conjuntos de dados gerados no processo de investigação e como e quando serão publicados. A gestão de dados deve ser efetuada segundo os princípios FAIR, tornando os dados fáceis de encontrar, acessíveis, interoperáveis e reutilizáveis (Comissão Europeia: Diretoria Geral para a Investigação e Inovação, 2016c). A implementação dos princípios FAIR na gestão e abertura de dados de investigação poderá potenciar também a partilha e por conseguinte, a citação de dados.

Em termos de dados públicos, a EU implementou uma nova Diretiva (2013/37/EU), que requer que a informação pública seja disponibilizada “through open and machine-readable formats and together with their metadata, at the best level of precision and granularity, in a format that ensures interoperability”, incluindo agora entidades como bibliotecas, arquivos e museus na esfera da abertura de dados (EUR-Lex, 2013:3). Esta diretiva tornou obrigatória a disponibilização de informação proveniente das entidades públicas para reutilização, ao custo de reprodução, de preferência num portal e sob licenças de formatos digitais.

A EU criou também o Open Data Maturity para avaliar o desenvolvimento de políticas nacionais e portais na Europa, usando a informação coletada para elaborar um índice de maturidade das políticas de Open Data baseado na preparação política e na maturidade do portal de dados (Carrara, Nieuwenhuis & Vollers, 2016). Em termos

de preparação política é avaliada a presença de uma política de Dados Abertos específica, normas de licenciamento, extensão da coordenação a nível nacional, uso de dados e impacto político, social e económico. A maturidade dos portais é avaliada de acordo com a sua usabilidade, a reutilização dos dados alojados no portal e inclusão de dados de diferentes domínios. Segundo a informação apurada no relatório, apesar da melhoria do número de políticas de Dados Abertos, no uso de dados e impacto da abertura de dados, a maioria dos países europeus ainda não atingiu a maturidade por completo. Os países necessitam de elevar o reconhecimento político sobre os Dados Abertos, aumentar os processos automatizados nos seus portais para aumentar a usabilidade e reutilização dos dados e também organizar mais eventos e treinos para apoiar as iniciativas nacionais e locais. O relatório do Open Data Maturity conclui que os países europeus estão a enfrentar barreiras legais, financeiras, de qualidade de dados e de políticas, sublinhando que os políticos não conhecem os benefícios da abertura de dados e que, por isso, os Dados Abertos não são estabelecidos como uma prioridade política (Carrara, Nieuwenhuis & Vollers, 2016).

Em Portugal, as políticas de abertura de dados de investigação e do setor público apresentam estados de maturidade diferentes. Apenas em 2014 foi definida uma política de AA aos dados de investigação, através da FCT enquanto entidade financiadora de investigação. Contudo, esta política “não está alinhada com os requisitos relativamente aos dados de investigação do H2020” (Comissão Executiva GT-PNCA, 2016), pois apenas recomenda a abertura de dados sem, no entanto, mencionar onde os depositar, o tipo de licenças a utilizar, que tipo de metadados inserir, que formatos utilizar.

A abertura de dados de investigação em Portugal foi enquadrada no âmbito da definição de uma política de CA, tendo sido delineadas algumas metas a atingir até 2018 pelo Ministério da Ciência, Tecnologia e Ensino Superior (2016): publicação de 100% dos dados criados com financiamento público num repositório de AA, verificação contínua da publicação de dados, eliminação do duplo financiamento para colocar publicações e dados em AA, oferta regular de formação relativa a publicação de dados em AA, propriedade intelectual e proteção de dados.

A Resolução do Conselho de Ministros nº 21/2016, de 11 de abril vem finalmente incluir o AA aos dados de investigação na política científica. No âmbito da definição de uma política de Ciência Aberta, é incluída a obrigatoriedade do AA aos “dados científicos resultantes de investigação financiada por fundos públicos” e sua respetiva preservação, “por forma a permitir a sua reutilização e o acesso continuado” (Resolução do Conselho de Ministros nº 21/2016, 2016). O interesse na abertura de dados de investigação tem crescido no discurso político e científico, verificando-se já a organização de uma conferência de dados de investigação e CA em 2016 e de dois fóruns de gestão de dados de investigação, o último já em 2017.

O primeiro relatório da Comissão Executiva do Grupo de Trabalho para a Política Nacional de Ciência Aberta constatou que, em termos de gestão e partilha de dados, existem algumas iniciativas e experiências de “âmbito disciplinar, nomeadamente no contexto de projetos”, sendo expectável que o Roteiro Nacional de Infraestruturas de Investigação de Interesse Estratégico traga “progressos significativos” (Comissão Executiva GT-PNCA, 2016:7). O mesmo relatório sublinha que ao nível das instituições de investigação e de ensino superior, as iniciativas de partilha e gestão de dados são quase inexistentes. De facto, apenas se regista a existência do projeto TAIL, lançado na Faculdade de Engenharia da Universidade do Porto, com o objetivo de otimizar o processo de gestão de dados de investigação, desde a sua produção até ao depósito e partilha. Este projeto, a decorrer de 2016 até 2019, “vai construir um portfólio de exemplos de gestão de dados em diversos domínios que poderão ser usados pelos investigadores para avaliar o esforço requerido e as compensações a obter com esta atividade” (Ribeiro, et al., 2016:256-257). Os resultados serão avaliados através da utilização da plataforma Dendro “designed to help researchers use concepts from domain-specific ontologies to collaboratively describe and share *datasets* within their groups” (Universidade do Porto, Faculdade de Engenharia, 2017).

O ecossistema dos dados de investigação em Portugal é concentrado a nível disciplinar e inferior em número de e-infraestruturas relativamente aos dados públicos. Regista-se a existência do SARDC com já cerca de seis anos de existência e que apresenta apenas 2007 *datasets* em maio de 2017. Depois da sua criação e da colocação em massa de conjuntos de dados, verifica-se que o depósito de *datasets* decresceu de forma extrema. Os conjuntos disponibilizados são provenientes de apenas duas instituições: o APIS com 37 e o IPCB com 1970 conjuntos de dados. Embora exista um espaço atribuído a mais duas instituições, nenhuma dessas entidades publicou, até agora, dados através desta plataforma. O SARDC é limitado em termos de funcionalidades, permitindo apenas a busca por autor, data, título, assunto e tipo de documento. Não é possível perceber a legalidade da reutilização de dados pois não é referido qualquer tipo de licenciamento. A qualidade dos dados apresentados pelo SARDC é uma incógnita pela falta de indicação do uso de formatos abertos e também de um esquema padrão de metadados. A nível das instituições de investigação, apenas o IPCB possui um repositório de Dados Abertos.

A tendência de abertura de dados de investigação parece estar ligada à cultura disciplinar, com os repositórios de dados científicos existentes em Portugal a conotarem-se no domínio das Ciências da Vida. O re3data indexa apenas 2 repositórios de dados científicos portugueses, o INTEGRALL e o KiMoSys, sendo que ambos pertencem ao ramo da biologia (DataCite, 2017a). O Biosharing.org (2017) indexou cinco bases de dados científicas com origem portuguesa, havendo participação nacional em outras duas bases de dados internacionais. Estas pertencem todas ao domínio das Ciências da Vida e da Biologia com exceção de uma, conotada com as

Ciências da Terra. A GBIF possui um nó português que aloja 23 conjuntos de dados que correspondem a 448136 ocorrências (Portal de Dados de Biodiversidade de Portugal, 2016). A rede nacional de informação biológica - BioData.pt fornece acesso a duas bases de dados disciplinares, uma ontologia para descrever experiências biológicas baseadas em plantas, quatro ferramentas de *software* e cinco conjuntos de dados.

Em suma, é necessário desenvolver por completo a política de abertura de dados científicos, clarificando questões como o local de depósito, licenças, metadados e formatos a utilizar. Ao nível do ecossistema é necessário melhorar a visibilidade do SARDC, abrir totalmente os dados, isto é, de forma legal e técnica e construir mais repositórios de dados. Em termos de impacto, é imperiosa uma mudança cultural que passe pelo treino dos investigadores para a gestão e partilha de dados, pela definição de recompensas e incentivos para os investigadores que partilhem dados e também pela criação de mais iniciativas de promoção da abertura de dados.

A evolução da conjuntura de acesso aos documentos administrativos do setor público é derivada da imposição das Diretivas Europeias, a que Portugal, como membro da EU, está obrigado. A Lei 46/2007 de 24 de Agosto virá transpor para a ordem jurídica nacional a Diretiva n.º 2003/98/CE, legislando a obrigação das entidades públicas publicar documentos administrativos, de forma eletrónica, para que possam ser acedidos e reutilizados. Foi também criada a CADA para supervisionar o cumprimento das disposições legais referentes ao acesso à informação administrativa” (Comissão de Acesso aos Documentos Administrativos [CADA]).

A Lei n.º 26/2016 de 22 de Agosto revoga e transpõe a novamente a Diretiva 2003/98/EC. O artigo 3º mandata a abertura dos dados ambientais à comunidade, incluindo-os na definição de documento administrativo, devendo a sua publicação ser feita num formato aberto e legível por máquina. O artigo 23º apenas prevê a reutilização dos dados para fins educativos e para fins de investigação e desenvolvimento, deixando de lado a reutilização comercial. Contudo, o dados.gov atribuiu aos dados que aloja a licença CC BY 3.0, o que não impede a reutilização comercial dos dados. Perante este contexto é impossível não questionar a razão de não ter sido transposta a nova Diretiva 2013/37/EU que facilita a reutilização de dados ao tornar obrigatória a disponibilização de informação proveniente das entidades públicas para reutilização, ao custo de reprodução, de preferência num portal e sob licenças de formatos digitais. Um desenvolvimento interessante seria o de criar um legislação apropriada para a abertura de dados governamentais, que fosse estabelecida tendo em conta as necessidades dos utilizadores e os objetivos governamentais, em vez da simples transposição da Diretiva da União Europeia.

Ao nível local, apenas Lisboa e Porto parecem ter definido uma política de abertura de dados, com estratégias diferenciadas. Lisboa aposta na publicação de dados municipais que promovam o envolvimento dos cidadãos, aumentem a transparência e

o desenvolvimento económico, além da melhoria da eficiência dos serviços, devendo os dados ser atualizados, legíveis por máquina, com licença aberta que permita a reutilização comercial (Lisboa. Câmara Municipal, 2016). Os dados disponibilizados no portal “são produzidos no âmbito de sistemas de informação” que pertencem a diversos organismos, registando-se já diversas aplicações que utilizam esses dados para criar valor para o cidadão e visitante de Lisboa na área do turismo, ambiente urbanismo (Soares, Baptista e Lima, 2015:18). Já o Porto parece concentrado na publicação de dados que contribuam para a “competitividade do Município do Porto e da região para visitar, para residir, captar e realizar negócios no mercado nacional e internacional” (Porto. Câmara Municipal). De acordo com a opinião de Soares, Baptista e Lima (2015:18), a abertura de dados é centrada na “dimensão turismo e negligência, pelo menos aparentemente, as restantes”. Esta conclusão pode ser reforçada com a análise do portal de dados do Porto, onde apenas constam cinco conjuntos de dados, sendo estes pertencentes ao turismo ou relacionados.

O ecossistema de abertura de dados tem aumentado dentro e fora do espectro político. Além do portal nacional de dados públicos, um grande número de organizações públicas tem criado os seus portais de dados e/ou tem publicado os mesmos no dados.gov. Por outro lado, existem vários portais de dados com origem na sociedade civil, com o selo da Transparência Hackday Portugal. Em Portugal têm existido diversas iniciativas a este nível, *hackathons* como o Date with Data, com o intuito de alertar, informar e formar a comunidade para este novo paradigma na disponibilização da informação conferências de Dados Abertos.

O Porto, Lisboa, Fundão, Palmela, Penela e Águeda estão entre as cidades que integram o Open and Agile Smart Cities (OASC) (2017), uma iniciativa com o objetivo de desenvolver mercados para as cidades inteligentes baseados nas necessidades das cidades e das comunidades. O Porto é destaque nesta iniciativa pela utilização da tecnologia FIWARE, que ajuda a melhorar o uso de dados abertos e a partilha conhecimentos e tecnologias entre cidades.

A análise feita por Soares, Baptista e Lima (2015) sobre a abertura de dados em quatro cidades do norte do país, Braga, Barcelos, Guimarães e Vila Nova de Famalicão, concluiu que a disponibilização de dados ao nível local é feita apenas ao nível de “contratos municipais, orçamento anual, gastos, resultados eleitorais, qualidade do ar, horários de transporte, orçamento participativo e lista de empresas”.

A avaliação da transparência governativa local pode também ser obtida através do Portal da Transparência Municipal, uma ferramenta que permite avaliar vários indicadores da governação dos municípios portugueses com o recurso à disponibilização de dados estatísticos. Os indicadores abordam a gestão financeira e administrativa, decisões fiscais, dinâmica económica do município, serviços municipais e participação eleitoral autárquica (Portugal. Governo, 2017).

Utilizando a estrutura de avaliação de políticas de Dados Abertos de Zuiderwijk e Janssen (2014), é possível tirar algumas ilações sobre o contexto, conteúdo e indicadores de desempenho da política de abertura de dados do setor público, bem como sobre o valor público alcançado. A nível de contexto da política de Dados Abertos, a Lei nº 26/2016 de 22 de Agosto transpõe a Diretiva 2003/98/CE, tendo como motivação para a sua implementação é sobretudo económica e política via EU. A política também pretende facilitar o acesso e reutilização da informação por parte dos cidadãos e garantir a transparência da atividade administrativa. A abertura de dados do setor público é garantida pela CADA, responsável pelo cumprimento da lei e a AMA que fornece a e-infraestrutura e apoio técnico. Também os custos de abertura de dados governamentais parecem estar assegurados num plano de racionalização de custos da Administração Pública (OECD, 2015).

Em termos de conteúdo, a política de Dados Abertos enuncia na sua constituição alguns princípios de abertura bastante importantes, como o acesso livre e universal, a acessibilidade, a interoperabilidade, a qualidade, a integridade e a autenticidade dos dados publicados e ainda a sua identificação e localização" (art. 2.º). A publicação de dados deve ser feita num formato aberto e legível por máquina (art. 3.º).

Relativamente ao contexto, o acesso é gratuito e de fácil compreensão, sendo que todos os dados no dados.gov estão disponíveis para *download* sem necessidade de pedido ou registo. É aplicada a licença CC BY 3.0 a todos os *datasets* alojados no portal, permitindo assim a sua reutilização com apenas a restrição de mencionar a fonte dos dados. Todos os dados no dados.gov estão disponíveis em formatos abertos podendo ser classificados com 3 estrelas, segundo o conceito de dados abertos conectados de Tim Berners-Lee (2010).

Contudo, não existe esquema padrão de metadados nem existe verificação dos metadados antes da publicação de cada *dataset* (Lapa, 2017). Não são conhecidas métricas de avaliação, nem multas ou recompensas no ecossistema português de Dados Abertos. A última medição da qualidade dos dados data de 2014, onde foram encontrados alguns problemas que não foram totalmente resolvidos junto dos fornecedores. A última grande atividade de promoção de reutilização de dados do setor público ocorreu em 2014-2015, não se conhecendo para já novas atividades previstas (Lapa, 2017).

Em termos de disponibilidade de dados, Portugal apresenta 802 conjuntos de dados no portal nacional, de temática diversa, mas com maior incidência na Economia e Finanças. Mas se considerarmos os dados alojado em outros portais públicos e não que ainda não constam no portal nacional, a contabilidade é bastante mais elevada.

Os indicadores de desempenho são modestos. Apesar do espetro alargado de destinatários da abertura de dados, registam-se apenas 2 aplicações elaboradas a

partir de Dados Abertos disponibilizados pela AMA e Direcção-Geral das Autarquias Locais (Lapa, 2017). Não é por isso surpreendente o fraco impacto político, económico e social apontado pelo Open Data Barometer e pelo Open Data Maturity. O estudo de Gomes e Soares (2014) conclui que a transparência é o único benefício real.

Em termos de valor público alcançado apenas conseguimos identificar a transparência governativa e disponibilidade de dados. Mesmo que a política mandate a abertura de dados mediante como a interoperabilidade e a qualidade, se não existe verificação nem medição, quem pode assegurar que estes estão a ser cumpridos? A disponibilidade de dados existe, o que não se verifica é grande reutilização dos dados públicos. Poderá esta falta de verificação da qualidade dos dados estar a interferir com a sua reutilização? Faltam claramente indicadores de reutilização de dados do setor público e estudos de impacto para uma avaliação concreta do valor alcançado.

Gomes e Soares (2014) apenas atribuem nota positiva ao dados.gov no cumprimento dos princípios de abertura. A atividade dos visitantes, o desenvolvimento de aplicações, a atividade geradora de novo conhecimento e a utilização de recursos do portal são classificadas como pobres. Neste indicador apenas a atividade dos fornecedores no portal é satisfatória. Em termos de benefícios observáveis da abertura de dados governamentais, o estudo apenas regista bons níveis de transparência governativa, enquanto a participação pública, a inovação e o empreendedorismo e a eficiência governativas não revelam impactos satisfatórios.

O cruzamento da informação obtida através das várias iniciativas de avaliação da abertura de dados permite inferir algumas conclusões e também revelar algumas disparidades. O Open Data Barometer e o Open Data Maturity apontam o baixo impacto económico e social da abertura de dados em Portugal. Gomes e Soares (2014) concluem que o problema em Portugal não é quantidade de dados disponibilizados mas sim a quantidade de dados que são reutilizados e como são reutilizados.

O Public Sector Information Scoreboard e o Open Data Maturity afirmam que Portugal não possui uma política de Dados Abertos, o que, como vimos, não corresponde à verdade. Entre as várias iniciativas de avaliação do estado de abertura de dados, apenas três publicaram informação relativa a 2016. Apesar da equivalência de ranking e taxa de abertura entre duas das iniciativas, é redundante inferir qualquer conclusão, dado que, estas não medem o mesmo número parâmetros, nem estes coincidem na maioria dos casos. O Open Data Inventory refere-se a dados avaliados apenas do portal do INE para calcular a sua taxa de abertura, pelo que, também deve ser relativizado.

Como pontos positivos são apontados a abertura de dados de diferentes domínios, o fato de todos os dados disponibilizados no portal estarem num formato processável

por máquina (Carrara, Nieuwenhuis & Vollers, 2016), ao contrário do que afirma a OECD (2015b) e o Open Data Inventory.

Todos os índices de avaliação da maturidade do estado das iniciativas e políticas de Dados Abertos são relativos a dados públicos, não tendo sido encontrada nenhuma estrutura de avaliação das políticas de abertura de dados de investigação.

Tendo em conta o Quadro 5, é possível verificar a qualidade da abertura dos dados dos 60 portais de dados portugueses encontrados na web e inferir algumas conclusões. Em termos de portais de dados de investigação foram encontrados apenas 11, com os restantes 49 a conter dados do setor público. Os portais de dados estatísticos são os grandes dominadores do panorama nacional, com 13 sítios web, seguidos dos portais de dados económicos com 6. Dos 60 portais analisados apenas 33 disponibilizam dados legíveis por máquina, sendo que alguns portais não o fazem totalmente. Apenas 16 portais indicam usar uma licença aberta que permita a livre reutilização dos dados e apenas 3 fornecem a indicação do padrão de metadados usado. Denota-se que alguns dos portais parecem usar um esquema próprio de descrição de *datasets*, no entanto, estes não parecem corresponder aos padrões de metadados mais reconhecidos. Existem 17 portais que não utilizam formatos abertos ou não indicam sequer o tipo de formato. É possível já encontrar já também 17 portais que fornecem dados em formatos estruturados.

Em suma, é necessário transpor a Diretiva da EU de 2013 e criar mais políticas municipais e institucionais, mandar a total abertura legal e técnica de dados e a sua verificação da qualidade. O nível de reutilização de dados e as atividades de treino e envolvimento dos cidadãos devem ser aumentados.

O sucesso das políticas de Dados Abertos depende, como vimos, da criação de um ecossistema sustentável e também do impacto e reutilização que estes tipos de informação possam induzir. O impacto dos Dados abertos pode ser medido através da citação de dados ou através de indicadores altimétricos, que permitem uma observação mais clara e detalhada dos possíveis impactos dos conjuntos de dados disponibilizados em AA, sejam eles provenientes do setor público ou das entidades de investigação.

Referências Bibliográficas

Agência para a Modernização Administrativa (AMA) (2016a). *Guia Dados Abertos*. Disponível na WWW: http://www.dados.gov.pt/media/46120/guia_dados_abertos_ama.pdf

Agência para a Modernização Administrativa (AMA) (2016b). *Dados.gov*. [Consultado a 11-05-2017]. Disponível na WWW: <http://www.dados.gov.pt/>

Agência Portuguesa do Ambiente (2017). *SNIAmb - Sistema Nacional de Informação de Ambiente*. [Consultado a 12-05-2017]. Disponível na WWW: <http://sniamb.apambiente.pt/Home/Default.htm>

Arzberger, P., et al. (2004). An international framework to promote access to data. *Science*, 303(5665), 1777-1778. Disponível na WWW: <http://science.sciencemag.org/content/303/5665/1777.full>

Asociación Española de Normalización y Certificación (2015). *UNE 178301: Ciudades Inteligentes, Datos Abiertos (Open Data)*. Madrid: AENOR.

Beagrie, C. & Houghton, J. (2014). The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres. Disponível na WWW: http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature*, 489(7415), 179. Disponível na WWW: <https://www.nature.com/news/predatory-publishers-are-corrupting-open-access-1.11385>

Bell, G. (2009). Foreword. In Hey, T., Tansley, S. & Tolle, K. (eds.) (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. XI-XV). Microsoft Research. Disponível na WWW: http://www.astro.caltech.edu/~george/aybi199/4th_paradigm_book_complete_lr.pdf

Berners-Lee, Tim (2010). *Is your Linked Open Data 5 Star?*. Disponível na WWW: <https://www.w3.org/DesignIssues/LinkedData.html>

Big Data Europe (2016). *Big Data Europe: Empowering Communities with Data Technologies*. [Consultado a 22-11-2016]. Disponível na WWW: <https://www.big-data-europe.eu/>

BioData.pt. *Recursos*. [Consultado a 18-05-2017]. Disponível na WWW: <http://www.biodata.pt/resources.html>

Biosharing.org (2017) *Policies*. [Consultado a 22-05-2017]. Disponível na WWW: <https://biosharing.org/policies/>

Björk, B.-C., Laakso, M., Welling, P. & Paetau, P. (2014). Anatomy of green open access. *Journal of the American Society for Information Science and Technology*, 65: 237–250. Disponível na WWW: <http://onlinelibrary.wiley.com/wol1/doi/10.1002/asi.22963/full>

BOAI - Budapest Open Access Initiative (2002). Disponível na WWW: <http://www.budapestopenaccessinitiative.org/>

BOAI - Budapest Open Access Initiative (2012). *Ten years on from the Budapest Open Access Initiative: setting the default to open*. Disponível na WWW: <http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

Borges, M. M. (2006). *A esfera: comunicação académica e novos média* (Tese de Doutoramento). Faculdade de Letras, Universidade de Coimbra, Portugal. Disponível na WWW: https://estudogeral.sib.uc.pt/jspui/bitstream/10316/8557/5/MMBorges_tese.pdf

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press

Bueno de la Fuente, G. (2016). *What is Open Science? Introduction*. Disponível na WWW: <https://www.fosteropenscience.eu/content/what-open-science-introduction>

Cabinet Office (2012). *Open Data White Paper: Unleashing the Potential*. Disponível na WWW: https://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf

Cabinet Office (2013). *G8 Open Data Charter and Technical Annex*. Disponível na WWW: <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>

Cambridge University Press (2017). *Cambridge Dictionary*. Disponível na WWW: <http://dictionary.cambridge.org/dictionary/english/policy>

Carrara, W., Chan, W. S., Fischer, S., Steenbergen, E. (2015). *Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources*. Disponível na WWW: https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

Carrara, W., Fischer, S., Oudkerk, F., S. Steenbergen, E. & Tinholt, D. (2015). *Analytical Report 1: Digital Transformations and Open Data*. Disponível na WWW: https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n1_-_digital_transformation.pdf

Carrara, W., Fischer, S. Steenbergen, E. (2015). *Analytical Report 2: E-skills and Open Data*. Disponível na WWW: https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n2_-_e-skills.pdf

Carrara, W., Nieuwenhuis, M. & Vollers, H. (2016). *Open Data Maturity in Europe 2016: Insights into the European state of play*. Disponível na WWW: https://www.capgemini-consulting.com/resource-file-access/resource/pdf/open_data_maturity_in_europe_2016_final_v1.0_0.pdf

Carrara, W., Oudkerk, F., Steenbergen, E., Tinholt, D. (2016a). *European Data Portal: Open Data Goldbook for Data Managers and Data Holders*. Capgemini Consulting. Disponível na WWW: <https://www.europeandataportal.eu/sites/default/files/goldbook.pdf>

Carrara, W., Engbers, W., Nieuwenhuis, M. & Steenberg, E. (2016b). *Analytical report n°4: Open Data in Cities*. Disponível na WWW: https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n4_-_open_data_in_cities_v1.0_final.pdf

Caruso, J., Nicol, A. & Archambault, É. (2013). Open Access Strategies in the European Research Area, (August), 1–46. Retrieved from http://europa.eu/rapid/press-release_IP-13-786_en.htm?locale=enhttp://www.science-metrix.com/pdf/SM_EC_OA_Policies.pdf

Carvalho, J., Furtado, F. & Príncipe, P. (2017). WP4 – D25 – Kit sobre dados de investigação. RCAAP. Disponível na WWW: <http://projecto.rcaap.pt/index.php/lang-pt/consultar-recursos-de-apoio/remository?func=finishdown&id=395>

Christensen-Dalsgaard, B., et al. (2012). *Ten recommendations for libraries to get started with research data management: Final report of the LIBER working group on E-Science / Research Data Management*. Disponível na WWW: <http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf>

CKAN. CKAN: *The open source data portal software*. Disponível na WWW: <http://ckan.org/>

Colpaert, P., Sarah, J., Peter, M., Mannens, E., Van de Walle, R. (2013). The 5 star of open data portals. *7th internacional conference on methodologies, Technologies and tools enabling e-Government (MeTTeG)*, 61-67. Disponível na WWW: <http://pieter.pm/5stardataportals/5starsofopendataportals.pdf>

Comissão de Acesso aos Documentos Administrativos (CADA). *Quem somos*. Disponível na WWW: <http://www.cada.pt/>

Comissão Europeia (2011). Open data: an engine for innovation, growth and transparent governance. Communication from the Commission to the European Parliament, the Council, the European economic and social committee and the committee of the regions (COM [2011] 882). Disponível na WWW: http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2011/0882/COM_COM%282011%290882_EN.pdf

Comissão Europeia (2012). COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES - Melhorar o acesso à informação científica: rentabilizar o investimento público em investigação. *COM (2012) 401 final*. Disponível na WWW: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2012:0401:FIN>

Comissão Europeia (2013). *EU implementation of the G8 Open Data Charter*. Disponível na WWW: <https://ec.europa.eu/digital-single-market/news/eu-implementation-g8-open-data-charter>

Comissão Europeia (2014). COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES - Para

uma economia dos dados próspera. COM(2014) 442 final. Disponível na WWW: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2014:0442:FIN>

Comissão Europeia (2016a). Open innovation, Open Science, open to the world. A vision for Europe. Brussels: European Commission, Directorate-General for Research and Innovation. Disponível na WWW: <http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbKI0416263/>

Comissão Europeia (2016b). Research & Innovation News Alert: Commission proposes copyright exception for researchers. Disponível na WWW: <http://ec.europa.eu/research/index.cfm?pg=newsalert&year=2016&na=na-140916>.

Comissão Europeia (2016c). COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES – Iniciativa Europeia para a Nuvem – Construir uma economia de dados e conhecimento competitiva na Europa. COM(2016) 178 final. Disponível na WWW: <http://eur-lex.europa.eu/legal-content/PT/ALL/?uri=CELEX:52016DC0178>

Comissão Europeia (2016d). Europa líder mundial da economia baseada nos dados, graças à iniciativa europeia para a computação em nuvem. *Comissão Europeia: comunicado de imprensa*. Disponível na WWW: http://europa.eu/rapid/press-release_IP-16-1408_pt.htm

Comissão Europeia (2017a). Digital Single Market: Digital Economy & Society. Disponível na WWW: <https://ec.europa.eu/digital-single-market/en/what-big-data-can-do-you>

Comissão Europeia (2017b). H2020 Programme: AGA – Annotated Model Grant Agreement. Disponível na WWW: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

Comissão Europeia (2017c). COMUNICAÇÃO DA COMISSÃO AO PARLAMENTO EUROPEU, AO CONSELHO, AO COMITÉ ECONÓMICO E SOCIAL EUROPEU E AO COMITÉ DAS REGIÕES - Construir uma economia de dados próspera. COM(2017) 9 final. Disponível na WWW: <http://ec.europa.eu/transparency/regdoc/rep/1/2017/PT/COM-2017-9-F1-PT-MAIN-PART-1.PDF>

Comissão Europeia: Grupo de peritos de alto nível em dados de investigação (2010). Riding the wave: how Europe can gain from the rising tide of scientific data. Disponível na WWW: http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204

Comissão Europeia: Diretoria Geral para a Investigação e Inovação (2016a). H2020 Programme: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Disponível na WWW: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Comissão Europeia: Diretoria Geral para a Investigação e Inovação (2016b). Realising the European Open Science Cloud: First report and recommendations Group, of the Commission

High Level Expert Cloud, on the European Open Science. Luxemburgo: Publications Office of the European Union. Disponível na WWW: https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

Comissão Europeia: Diretoria Geral para a Investigação e Inovação (2016c). H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. Disponível na WWW: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Comissão Europeia: Diretoria Geral para a Investigação e Inovação (2017). Open Science Monitor. Disponível na WWW: <http://ec.europa.eu/research/openscience/index.cfm?pg=about§ion=monitor>

Comissão Executiva Grupo de Trabalho para a Política Nacional de Ciência Aberta (Comissão Executiva GT-PNCA) (2016). Primeiro Relatório da Comissão Executiva Grupo de Trabalho Para a Política Nacional de Ciência Aberta. Disponível na WWW: http://media.wix.com/ugd/a8bd7c_1ca622bff7f34abbad228ac94e1eda16.pdf

Concordat on Open Research Data (2016). Disponível na WWW: <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/>

Croll, A. (2012). The feedback economy. In Planning for big data (pp. 1-8). Sebastopol: O'Reilly Media.

Cybermetrics Lab (2017). *Ranking web of repositories*. [Consultado a 02-02-2017]. Disponível na WWW: <http://repositories.webometrics.info/en>

Dallmeier-Tiessen, et al. (2011). Highlights from the SOAP project survey. What Scientists Think about Open Access Publishing. *arXiv:1101.5260 [Cs]*, 14. <https://arxiv.org/abs/1101.5260>

DataCite (2017a). *re3data: registry of research data repositories*. [Consultado a 02-05-2017]. Disponível na WWW: <http://www.re3data.org/>

DataCite (2017b). *Search*. [Consultado a 02-05-2017]. Disponível na WWW: <https://search.datacite.org/>

Data Documentation Initiative (DDI) (2016). *Document, Discover and Interoperate*. Disponível na WWW: <http://www.ddialliance.org/>

DataHub. *Datasets*. [Consultado a 03-01-2017]. Disponível na WWW: <https://datahub.io/dataset?q=portugal&page=1>

DataCite (2017). *Search*. [Consultado a 18-05-2017]. Disponível na WWW: <https://search.datacite.org/>

Davies, T. (2012). *5-Stars of Open Data Engagement?*. Disponível na WWW: <http://www.timdavies.org.uk/2012/01/21/5-stars-of-open-data-engagement/>

Davies, T. (2014). Open data policies and practice: an international comparison. *Paper for European Consortium for Political Research Panel P356 –The Impacts of Open Data*. Disponível na WWW: <https://ecpr.eu/Filestore/PaperProposal/d591e267-cbee-4d5d-b699-7d0bda633e2e.pdf>

Dawes, S., Vidasova, L. & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15-27. Disponível na WWW: <http://www.sciencedirect.com/science/article/pii/S0740624X1630003X>

Declaração de Bethesda sobre a Publicação em Acesso Aberto (2003). Disponível na WWW: <http://legacy.earlham.edu/~peters/fos/bethesda.htm>

Declaração de Berlim sobre Acesso Livre ao Conhecimento nas Ciências e Humanidades (2003). Disponível na WWW: <https://repositorium.sdum.uminho.pt/about/DeclaracaoBerlim.htm>

Decreto-Lei n.º 115/2013 de 7 de agosto do Ministério da Educação e Ciência. Diário da República, 1.ª série. N.º 151 (2013), 4749-4772. Disponível na WWW: <https://dre.pt/application/dir/pdf1sdip/2013/08/15100/0474904772.pdf>

Despacho n.º 15389/2016 de 12 de dezembro do Ministério das Finanças e da Secretaria de Estado da Cultura e Ciência, Tecnologia e Ensino Superior. Diário da República n.º 244/2016, Série II, N.º 244 (2016), 37155 – 37156. Disponível na WWW: <https://dre.pt/application/file/a/105580126>

Dietrich, D. (2013). *Open Government Data: principles, concepts & best practice*. Disponível na WWW: <http://workspace.unpan.org/sites/Internet/Documents/UNPAN95209.pdf>

Dietrich, D., Cherkasov, V., Susar, D. (2013). *Open Government Data for Citizen Engagement in Managing Development*. Disponível na WWW: <http://workspace.unpan.org/sites/Internet/Documents/OGDCE%20Toolkit%20v1.13-Feb2013.pdf>

Directory of Open Access Journals (DOAJ) (2017). *Search*. [Consultado a 27-05-2017]. Disponível na WWW: <https://doaj.org/>

Dodds, L., Newman, A. (2015). *Open Data Maturity Model ed. 1.0*. Disponível na WWW: <http://theodi.org/guides/maturity-model>

Donker, F. & van Loenen, B. (2017). How to assess the success of the open data ecosystem?. *International journal of digital earth*, 10(3), 284-306. Disponível na WWW: <http://www.tandfonline.com/doi/full/10.1080/17538947.2016.1224938>

Dumbill, E. (2012). What is Big Data. In *Planning for big data* (pp. 9-16). Sebastopol: O'Reilly Media. Disponível na WWW: <http://eecs.wsu.edu/~yinghui/mat/courses/fall%202015/resources/planning-for-big-data.pdf>

e-Infrastructure Reflection Group (e-IRG) (2016). *Long Tail of Data: e-IRG Task Force Report*. Disponível na WWW: <http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf>

ELIXIR Portugal (2015). *What is ELIXIR Portugal?*. Disponível na WWW: <http://elixir-portugal.org/>

ePSI Scoreboard (2013). ePSI platform scoreboard. Disponível na WWW: https://www.europeandataportal.eu/sites/default/files/overview_page_-_epsi_platform_scoreboard.pdf

Etalab (2011). *Licence Ouvert*. Disponível na WWW: <https://www.etalab.gouv.fr/en/licence-ouverte-open-licence>

EUDAT(a). *What is EUDAT?*. Disponível na WWW: <https://eudat.eu/what-eudat>

EUDAT(b). *B2FIND*. [Consultado a 22-05-2017]. Disponível na WWW: <http://b2find.eudat.eu/>

EUR-Lex (2003a). *Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information*. Official Journal of the European Union, 90-96. Disponível na WWW: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:en:PDF>

EUR-Lex (2003b). *Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC*. Official Journal of the European Union, 26-32. Disponível na WWW: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:041:0026:0032:EN:PDF>

EUR-Lex (2011). *COMMISSION DECISION of 12 December 2011 on the reuse of Commission documents (2011/833/EU)*. Official Journal of the European Union, 39-42. Disponível na WWW: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:330:0039:0042:EN:PDF>

EUR-Lex (2012). *COMMISSION RECOMMENDATION of 17 July 2012 on access to and preservation of scientific information (2012/417/EU)*. Official Journal of the European Union, 39-43. Disponível na WWW: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32012H0417>

EUR-Lex (2013). *Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information*. Official Journal of the European Union, 1-8. Disponível na WWW: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>

EUR-Lex (2016). *REGULAMENTO (UE) 2016/679 DO PARLAMENTO EUROPEU E DO CONSELHO de 27 de abril de 2016 relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados)*. Jornal Oficial da União Europeia, 1-88. Disponível na WWW: <http://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679&from=PT>

European Data Portal (2016). *Conjuntos de dados*. [Consultado a 12-05-2017]. Disponível na WWW: <https://www.europeandataportal.eu/data/pt/dataset>

European Research Area and Innovation Committee (ERAC) (2016). *ERAC opinion on open research data (ERAC 1202/16)*. Disponível na WWW: [http://www.earto.eu/fileadmin/content/Website/ERAC Opinion on Open Research Data.PDF](http://www.earto.eu/fileadmin/content/Website/ERAC%20Opinion%20on%20Open%20Research%20Data.PDF)

Erway, R., Horton, L., Nurnberger, A., Otsuji, R. & Rushing, A. (2016). *Building blocks: laying the foundation for a research data management program*. Dublin, Ohio: OCLC Research. Disponível na WWW: <https://www.oclc.org/content/dam/research/publications/2016/oclcresearch-data-management-building-blocks-2016.pdf>

Expert Advisory Group on Data Access (EAGDA) (2014). *Establishing Incentives and changing cultures to support data access*. Disponível na WWW: <https://wellcome.ac.uk/sites/default/files/establishing-incentives-and-changing-cultures-to-support-data-access-eagda-may14.pdf>

Executive Office of the President: Office of Management and Budget (2013). *MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: Open Data Policy - Managing Information as an Asset*. Disponível na WWW:

<https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>

Fan, W. & Bifet, A. (2012). Mining Bid Data: Current Status, and Forecast to the Future. *SIGKDDExplorations* 14(2), 1-5. Disponível na WWW: http://www.kdd.org/exploration_files/V14-02-01-Fan.pdf

Fecher, B. & Friesike, S. (2013). Open Science: One Term, Five Schools of Thought. In Bartling, S. & Friesike (Eds.), *Opening Science* (pp. 17-47). New York: Springer. Disponível na WWW: https://www.researchgate.net/publication/236607487_Open_Science_One_Term_Five_Schools_of_Thought

FIWARE (2015). Porto, a city that has become a real-time guide. Disponível na WWW: <https://www.fiware.org/2015/11/20/porto-a-city-that-has-become-a-real-time-guide/>

FORCE11: FAIR Data Publishing Group (2014). *The FAIR Data Principles*. Disponível na WWW: <https://www.force11.org/group/fairgroup/fairprinciples>

Forsström, P., & Haataja, J. (2016). *Open Science as an Instrument for Effective Research*. Disponível na WWW: <http://ojs.tsv.fi/index.php/signum/article/viewFile/58741/20303>

FOSTER (2016). *About FOSTER portal*. Disponível na WWW: <https://www.fosteropenscience.eu/about>

Fundação para a Ciência e Tecnologia [FCT] (2014). *Política sobre a Disponibilização de Dados e outros Resultados de Projetos de I&D Financiados Pela FCT*. Disponível na WWW: https://www.fct.pt/documentos/PoliticaAcessoAberto_Dados.pdf

G20's Anti-corruption Working Group (2014). G20 Anti-corruption Open Data Principles. Disponível na WWW: <http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf>

Garcia, M. (2015) New businesses around open data, smart cities and Fiware. European Public Sector Information Platform, Topic Report Nº 2015/04. Disponível na WWW: <http://www.investhorizon.eu/resources/documents/6/263232145-New-Businesses-Around-Open-Data-Smart-Cities-FIWARE.pdf>

Garfield, E. (2006). The History and Meaning of the Journal Impact Factor. *Journal of the American Medical Association (JAMA)*, 293, 90-93. Disponível na WWW: <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf>

GitHub (2017). Search. [Consultado a 18-05-2017]. Disponível na WWW: <https://github.com/search?utf8=%E2%9C%93&q=portugal&type=>

Gomes, A. & Soares, D. (2014). Open government initiatives in Europe: northern versus Southern countries analysis. In *ICEGOV2014, Outubro 27-30, Guimarães, Portugal*. Disponível na WWW: <http://hdl.handle.net/1822/37172>

Gonzalez-Zapata, F. & Heeks, R. (2015) The multiple meanings of open government data: understanding different stakeholders and their perspectives, *Government Information Quarterly*, 32(4), 441-452. Disponível na WWW: <https://doi.org/10.1016/j.giq.2015.09.001>

Granickas, K. (2013). Understanding the impact of releasing and re-using open government data. European Public Sector Information Platform, Topic Report N.º 2013 / 08. Disponible na WWW:

https://www.europeandataportal.eu/sites/default/files/library/201308_understanding_the_impact_of_releasing_and_re_using_open_data.pdf

Hahnel, M. & Hook, D. (2016). Open by default?. In Treadway, J., et al. *The state of open data: a selections of analyses and articles about open data, curated by Figshare* (pp. 3-6). Disponible na WWW: https://figshare.com/articles/The_State_of_Open_Data_Report/4036398/1

Harnad, S., et al. (2004). The Access/Impact Problem and the green and gold roads to Open Access. *Serials Review* 30(4). Disponible na WWW: <http://users.ecs.soton.ac.uk/harnad/Temp/impact.html>

Harnad, S. & Brody, T. (2004) Comparing the Impact of Open Access (OA) vs. Non-Open Access Articles in the Same Journals, *D-Lib Magazine*, 10(6). Disponible na WWW: <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

Hernández-Pérez, T. (2016). En la era de la web de los datos: primero datos abiertos, después datos masivos. *El profesional de la información*, 25(4), 517-525. Disponible na WWW: <https://doi.org/10.3145/epi.2016.jul.01>

Hey, T., Tansley, S. & Tolle, K. (2009). Jim Gray on eScience: A Transformed Scientific Method. In Hey, T., Tansley, S. & Tolle, K. (eds.) (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. 5-12). Microsoft Research. Disponible na WWW: http://www.astro.caltech.edu/~george/aybi199/4th_paradigm_book_complete_lr.pdf

Hogge, B. (2010). Open Data Study. Open Society Institute. Disponible na WWW: http://www.transparencyinitiative.org/wpcontent/uploads/2011/05/open_data_study_final.pdf

Houghton, J., & Swan, A. (2013). Planting the green seeds for a golden harvest: Comments and clarifications on "Going for Gold." *D-Lib Magazine*, 19, 1-2. Disponible na WWW: <http://doi.org/10.1045/january2013-houghton>

Huijboom, N., van den Broek, T. (2011). Open Data: an international comparison of strategies. *European Journal of ePractice*, 12, 4-16. Disponible na WWW: <https://joinup.ec.europa.eu/sites/default/files/76/a7/05/ePractice%20Journal-%20Vol.%2012-March%20April%202011.pdf>

International Council for Science (ICSU) (2014). *Open access to scientific data and literature and the assessment of research by metrics*. Disponible na WWW: <http://wiki.lib.sun.ac.za/images/b/b8/Icsu-report-on-open-access.pdf>

International Open Data Charter (2015). *Open Data Charter*. Disponible na WWW: <http://opendatacharter.net/>

International Workshop on Open Data for Science and Sustainability in Developing Countries (2014). *Data Sharing Principles in Developing Countries (The Nairobi Data Sharing Principles)*.

Disponível na WWW: <https://www.rd-alliance.org/sites/default/files/attachment/NairobiDataSharingPrinciples.pdf>

Janssen, M. (2011). The influence of the PSI directive on open government data: an overview of recent developments. *Government Information Quarterly*, 28, 446-456. Disponível na WWW: <https://doi.org/10.1016/j.giq.2011.01.004>

Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012) Benefits, adoption, barriers and myths of open data and open government. *Information Systems Management*, 4, 258-268. Disponível na WWW: <http://www.tandfonline.com/doi/abs/10.1080/10580530.2012.716740>

Jensen, K. (2012). Open Government Data and the right to information: opportunities and obstacles. *The Journal of Community Informatics* 8 (2). Disponível na WWW: <http://ci-journal.net/index.php/ciej/article/view/952>

Koski, H. (2015). The impact of open data – a preliminary study. Ministry of Finance publications, 15b/2015. Disponível na WWW: https://www.w3.org/2013/share-psi/wiki/images/6/67/Impact_of_Open_Data_in_the_Public_Sector_Koski_2015.pdf

Kramer, B. & Bosman, J. (2016). *Innovations in Scholarly Communication*. [Consultado a 24-02-2017]. Disponível na WWW: <https://101innovations.wordpress.com/>

Laakso, M. (2014). Green open access policies of scholarly journal publishers: A study of what, when, and where self-archiving is allowed. *Scientometrics*, 99(2), 475–494. Disponível na WWW: <http://doi.org/10.1007/s11192-013-1205-3>

Lämmerhirt, D. (2016). PASTEUR4OA Briefing Paper: Disciplinary differences in opening research data. Disponível na WWW: <http://www.pasteur4oa.eu/resources/209>

Laney, D. (2001). 3-D Data Management: Controlling DataVolume, Velocity and Variety. META Group. Disponível na WWW: <https://www.noexperiencenecessarybook.com/8VeXJ/3d-data-management-controlling-data-volume-velocity-and-variety.html>

Lei nº 46/2007 de 24 de Agosto. Regula o acesso aos documentos administrativos e a sua reutilização, revoga a Lei n.º 65/93, de 26 de Agosto, com a redação introduzida pelas Lei nº 8/95, de 29 de Março, e 94/99, de 16 de Julho, e transpõe para a ordem jurídica nacional a Diretiva n.º 2003/98/CE, do Parlamento e do Conselho, de 17 de Novembro, relativa à reutilização de informações do sector público. *Diário da República*, 1.ª série. N.º 163 (2007), 5680-5687. Disponível na WWW: https://www.fct.pt/arquivo/docs/Lei46_2007.pdf

Lei n.º 36/2011 de 21 de Junho. Estabelece a adopção de normas abertas nos sistemas informáticos do Estado. *Diário da República*, 1.ª série — N.º 118 (2011), 3599-3600. Disponível na WWW: <https://m6.ama.pt/docs/Lei362011-NormasAbertas.pdf>

Lei nº 26/2016 de 22 de Agosto. Aprova o regime de acesso à informação administrativa e ambiental e de reutilização dos documentos administrativos, transpondo a Diretiva 2003/4/CE, do Parlamento Europeu e do Conselho, de 28 de janeiro, e a Diretiva 2003/98/CE, do Parlamento Europeu e do Conselho, de 17 de novembro. *Diário da República*, 1.ª série. N.º 160 (2016), 2777-2788. Disponível na WWW: <https://dre.pt/application/conteudo/75177801>

Leonelli, S. (2016). Why open data now? Big data, knowledge production and the political economy of research. In Treadway, J., et al. *The state of open data: a selections of analyses and articles about open data, curated by Figshare* (pp. 7-11). Disponível na WWW: https://figshare.com/articles/The_State_of_Open_Data_Report/4036398/1

Leonelli, S., Spichtinger, D., and Prainsack, B. (2015) Sticks and carrots: encouraging open science at its source. *Geography and Environment*, 2: 12–16. Disponível na WWW: <http://onlinelibrary.wiley.com/enhanced/exportCitation/doi/10.1002/geo2.2>

Lisboa. Câmara Municipal (2016). *Smart Open Lisboa*. Disponível na WWW: <http://www.smartopenlisboa.com/>

Lisboa. Câmara Municipal (2017). *Lisboa Aberta*. [Consultado a 12-05-2017]. Disponível na WWW: <http://dados.cm-lisboa.pt/dataset>

Lynch, C. (2009). Jim Gray's Fourth Paradigm and the Construction of the Scientific Record. In Hey, T., Tansley, S. & Tolle, K. (eds.) (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. 177-183). Microsoft Research. Disponível na WWW: http://www.astro.caltech.edu/~george/aybi199/4th_paradigm_book_complete_lr.pdf

Machado, Jorge (2015). Open Data and Open Science. In Albagli, S., Maciel, M. & Abdo, A. (Eds.). *Open Science, Open Issues: 189-214*. Brasília: IBICT; Rio de Janeiro: Unirio. Disponível na WWW: http://livroaberto.ibict.br/bitstream/1/1061/1/Open%20Science%20open%20issues_Digital.pdf

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. (2011). *Big Dat: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. Disponível na WWW: <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>

Marr, B. (2014). *Big Data: The 5 Vs Everyone Must Know*. Disponível na WWW: <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know>

Nações Unidas (2015). *Addis Ababa Action Agenda of the Third International Conference on Financing for Development (Addis Ababa Action Agenda)*. Disponível na WWW: https://sustainabledevelopment.un.org/content/documents/2051AAAA_Outcome.pdf

National Institutes of Health (NIH) (2008). *NIH Guide Notice for Public Access*. Disponível na WWW: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html>

Nature Springer (2016). *Over 600 Springer Nature journals commit to new data sharing policies*. Disponível na WWW: <http://www.springernature.com/br/group/media/press-releases/over-600-springer-nature-journals-commit-to-new-data-sharing-policies/11111248>

OECD (2007). *Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publishing. Disponível na WWW: <http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>

OECD (2013). Open Government data. In *Government at a Glance 2013* (pp. 140-142). Paris: OECD Publishing. Disponível na WWW: http://dx.doi.org/10.1787/gov_glance-2013-48-en

OECD (2015a). Making Open Science a Reality. *OECD Science, Technology and Industry Policy Papers*, 25. Paris: OECD Publications. Disponível na WWW: <http://doi.org/10.1787/5jrs2f963zs1-en>

OECD (2015b). Assessing government initiatives on public sector information: A review of the OECD Council Recommendation. *OECD Digital Economy Papers*, 248. Paris: OECD Publishing. Disponível na WWW: <http://dx.doi.org/10.1787/5js04dr9l47j-en>

OECD (2016a) Open Government Data. Disponível na WWW: <http://www.oecd.org/gov/digital-government/open-government-data.htm>

OECD (2016b). *Observatory of Public Sector Innovation*. Disponível na WWW: <https://www.oecd.org/governance/observatory-public-sector-innovation/home/>

Olavsrud, T. (2013, Abril, 05). 10 intriguing real-world uses for big data. *Computer World*. Disponível na WWW: <http://www.computerworld.com/article/2473691/big-data/92712-10-Real-World-Big-Data-Deployments-That-Will-Change-Our-Lives.html>

Open Access Directory (2016). *Data repositories*. [Consultado a 22-02-2017]. Disponível na WWW: http://oad.simmons.edu/oadwiki/Data_repositories

OpenAIRE. *Research Data*. [Consultado a 12-05-2017]. Disponível na WWW: <https://www.openaire.eu/search/find/datasets?keyword=>

OpenAIRE (2017). *About OpenAIRE-Connect*. Disponível na WWW: <https://www.openaire.eu/connect>

Open and Agile Smart Cities (OASC) (2017). *Cities*. [Consultado a 17-05-2017]. Disponível na WWW: <http://www.oascities.org/list-of-cities/>

Open Data for Development Network (OD4D[a]). *About OD4D*. [Consultado a 16-02-2017]. Disponível na WWW: <http://od4d.com/about/>

Open Data for Development Network (OD4D[b]). *Open Data Impact Map*. [Consultado a 16-02-2017]. Disponível na WWW: <http://opendataimpactmap.org/index.html>

Open Data for Development Network (OD4D) (2016). Summary report and the Second Action Plan for International Collaborations: International Open Data Roadmap. International Open Data Conference 2016, 6-7 outubro. Madrid, Espanha. Disponível na WWW: <http://od4d.com/roadmap/assets/files/report-iodc-2016-web.pdf>

Open Data Institute (ODI) (2012). *About the ODI*. Disponível na WWW: <http://theodi.org/about>

Open Data Institute (ODI) (2015). Open Data Maturity Model: assessing your open data publishing and use. Disponível na WWW: <https://pt.scribd.com/document/260481608/ODI-Maturity-Model-Guide-Assessing-your-open-data-publishing-and-use>

Open Data Soft (2017). A comprehensive list of 2600+ Open Data portals around the world. [Consultado a 15-04-2017]. Disponível na WWW: <https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/>

Open Government Partnership (OGP) (2011). *Open Government Declaration*. Disponível na WWW: <https://www.opengovpartnership.org/open-government-declaration>

Open Government Partnership (OGP) (2015). *What is the Open Government Partnership?*. Disponível na WWW: <http://www.opengovpartnership.org/>

Open Government Partnership (OGP) (2016). PARIS DECLARATION: 4th Global Summit of the Open Government Partnership. Disponível na WWW: http://www.opengovpartnership.org/sites/default/files/OGP-Summit_PARIS-DECLARATION_EN.pdf

Open Knowledge International (OKI [a]). *Open Definition 2.1*. Disponível na WWW: <http://opendefinition.org/od/2.1/en/>

Open Knowledge International (OKI [b]). *Linked Open Vocabularies*. [Consultado a 31-01-2017]. Disponível na WWW: <http://lov.okfn.org/dataset/lov>

Open Knowledge International (OKI [c]). *Open Data Commons*. Disponível na WWW: <http://opendatacommons.org/>

Open Knowledge International (OKI [d]). *Datahub*. [Consultado a 18-02-2017]. Disponível na WWW: <https://datahub.io/>

Open Knowledge International (OKI) (2012a). *The Open Data Handbook*. Disponível na WWW: <http://opendatahandbook.org/guide/en/>

Open Knowledge International (OKI) (2012b). *What is open*. Disponível na WWW: <https://okfn.org/opendata/>

Open Knowledge International (OKI) (2014). *Portugal Open Data Census*. Disponível na WWW: <http://pt-city.census.okfn.org/year/2014>

Open Knowledge International (OKI) (2017). *Global Open Data Index*. Disponível na WWW: <http://index.okfn.org/place/>

Open Knowledge International: Working Group on Open Government Data (?). *Open Government Data*. Disponível na WWW: <http://opengovernmentdata.org/>

Open Science and Research Initiative (2014). *The Open Science and Research Handbook*. Disponível na WWW: <https://avointiede.fi/documents/14273/0/Open+Science+and+Research+Handbook+v.1.0/50316d5d-440b-4496-b039-2997663afff8>

Open Science Monitor (2015). *Open Science Monitor*. Disponível na WWW: <http://opendatamonitor.eu/>

Pangea (2017). *Search*. [Consultado a 18-05-2017]. Disponível na WWW: <https://www.pangaea.de/?q=Portugal>

Panton Principles (2010). *Panton Principles, Principles for open data in science*. Disponível na WWW: <http://pantonprinciples.org/>

Pasqueto, I., Sands, A. & Borgman, C. (2015, novembro). Exploring openness in data and science: What is “open”, to whom, when and why? *Apresentado em ASIST 2015, 6 - 10 Novembro*. St. Louis, MO, USA. Disponível na WWW: <http://onlinelibrary.wiley.com/doi/10.1002/pr2.2015.1450520100141/full>

Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015, outubro). Fostering Open Science to Research using a Taxonomy and an eLearning Portal. *Apresentado em I-KNOW '15: Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business, 21 - 22 Outubro 2015*, 1–8. Graz, Austria. Disponível na WWW: <http://doi.org/10.1145/2809563.2809571>

Pontika, N., Knoth, P., Cancellieri, M., & Pearce, S. (2015). Open Science Taxonomy. *Figshare*. Disponível na WWW: <https://dx.doi.org/10.6084/m9.figshare.1508606.v3>

Portal de Dados de Biodiversidade de Portugal (2016). *Dados*. [Consultado a 18-05-2017]. Disponível na WWW: <http://dados.gbif.pt/>

Portaria n.º 285/2015 de 15 de setembro do Ministério da Educação e Ciência. Diário da República, 1.ª série. N.º 180 (2015), 8011-8014. Disponível na WWW: <https://dre.pt/application/conteudo/70297426>

Porto. Câmara Municipal. *Dados Abertos: para uma cidade mais inteligente*. [Consultado a 22-05-2017]. Disponível na WWW: <http://dadosabertos.cm-porto.pt/about>

Portugal. Governo (2017). *Portal da Transparência Municipal*. [Consultado a 20-05-2017]. Disponível na WWW: <https://www.portalmunicipal.pt/home?locale=pt>

Portugal. Ministério da Ciência, Tecnologia e Ensino Superior, Ministério da Cultura (2016). *Diretório dos Repositórios Digitais*. Disponível na WWW: <http://repositorios-conhecimento.pt/>

Portugal. Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) (2016). *Ciência Aberta, Conhecimento para todos: Princípios orientadores*. Disponível na WWW: <http://www.portugal.gov.pt/media/18506199/20160210-mctes-ciencia-aberta.pdf>

Price, D., & Chaytor, S. (2015). Open Access: is a national licence the answer?. Disponível na WWW: <http://www.hepi.ac.uk/2015/03/30/open-access-national-licence-answer/?una-licencia-nacional-es-la-respuesta/>

Pröl, S., Meixner, K. & Rauber, A. (2016). Precise Data Identification Services for Long Tail Research Data. *Proceedings of International Conference on Digital Preservation, Bern, Switzerland*. Disponível na WWW: <https://www.rd-alliance.org/system/files/documents/iPRES2016-Proell.pdf>

Rans, J., Day, M., Duke, M., & Ball, A. (2013). Enabling the citation of datasets generated through public health research, (DCC), 41. Retrieved from <https://wellcome.ac.uk/sites/default/files/enabling-citation-of-datasets-through-public-health-research-phrdf-2013.PDF>

RECODE (2014) Policy recommendations for open access to research data. Disponível na WWW: http://recodeproject.eu/wp-content/uploads/2015/01/recode_guideline_en_web_version_full_FINAL.pdf

Registry of Open Access Repository Mandates and Policies (ROARMAP). *Browse*. [Consultado a 25-11-2016]. Disponível na WWW: <http://roarmap.eprints.org/view/country/620.html>

Reitano, S. (2013). The Benefits of Open Data. *Royal Roads University*. Disponível na WWW: <http://docplayer.net/589190-The-benefits-of-open-data.html>

Resolução do Conselho de Ministros n.º 91/2010. Diário da República, 1.ª série. N.º 225 (2010), 5288-5307. Disponível na WWW: <https://dre.pt/application/dir/pdf1s/2010/11/22500/0528805307.pdf>

Resolução do Conselho de Ministros n.º 21/2016. Diário da República, 1.ª série. N.º 70 (2016), 1191-1193. Disponível na WWW: <https://dre.pt/application/conteudo/74094659>

Research Data Alliance (RDA) (2015). 23 Things: Libraries for Research Data. Disponível na WWW: https://www.rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_Management.pdf

Ribeiro, C., et al. (2016). Projeto TAIL – Gestão de dados de investigação, da produção ao depósito e à partilha (resultados preliminares). *Cadernos BAD*, 2 (jul-dez), 256-264. Disponível na WWW: <http://www.bad.pt/publicacoes/index.php/cadernos/article/viewFile/1603/pdf>

Rouder, J. (2016). The what, why, and how of born-open data. *Behavior Research Methods*, 48(3), 1062-1069. Disponível na WWW: <http://link.springer.com/article/10.3758/s13428-015-0630-z>

Royal Society (2012). Science as an Open Enterprise. *The Royal Society Policy Centre Report*, 02/12. Disponível na WWW: <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>

Sasse, T., Smith, A., Broad, E., Tennison, J., Wells, P. & Atz, U. (2017) Recommendations for Open Data Portals: from Setup to sustainability. Disponível na WWW: https://www.europeandataportal.eu/sites/default/files/edp_s3wp4_sustainability_recommendations.pdf

Schiermeier, Q. (2015, Dezembro, 04). Pirate research-paper sites play hide-and-peek with publishers. *Nature: Internacional weekly journal of science*. Disponível na WWW: <http://www.nature.com/news/pirate-research-paper-sites-play-hide-and-peek-with-publishers-1.18876>

Science International (2015). Open Data in a Big Data World. Paris: International Council for Science (ICSU), International Social Science Council (ISSC), The World Academy of Sciences (TWAS), InterAcademy Partnership (IAP). Disponível na WWW: http://www.science-international.org/sites/default/files/reports/open-data-in-big-data-world_short_en.pdf

Science International (2016). *About Science International*. Disponível na WWW: <http://www.science-international.org/#about>

ScienceOpen (2017). *About ScienceOpen*. Disponível na WWW: <http://about.scienceopen.com/category/about-scienceopen/>

Sci-Hub. *About*. Disponível na WWW: <http://sci-hub.io/>

Shadbolt, N. (2016). Foreword. In Treadway, J., et al. *The state of open data: a selections of analyses and articles about open data, curated by Figshare* (pp. 2). Disponível na WWW: https://figshare.com/articles/The_State_of_Open_Data_Report/4036398/1

SHERPA/JULIET. *Research funders' open access policies*. [Consultado a 26-11-2016]. Disponível na WWW: <http://sherpa.ac.uk/juliet/index.php?la=en&mode=simple>

SHERPA/RoMEO. *Publisher copyright policies & self-archiving*. [Consultado a 26-11-2016]. Disponível na WWW: <http://sherpa.ac.uk/romeo/index.php?la=en&fIDnum=1109|1&mode=advanced>

Soares, N., Baptista, A. & Lima, A. (2015). Estudo Open Data Municípios Quadrilátero. Disponível na WWW: http://www.ccg.pt/wp-content/uploads/2015/10/OpendataQuadrilatero_Estudo.pdf

SPARC. *Open Data*. Disponível na WWW: <http://sparcopen.org/open-data/>

Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3) (2014). *What is SCOAP3?*. Disponível na WWW: <https://scoap3.org/what-is-scoap3/>

Statement on Data Sharing in Public Health Emergencies (2016). Disponível na WWW: <http://blogs.plos.org/plos/2016/02/statement-on-data-sharing-in-public-health-emergencies/>

Sturges, P., Bamkin, M., Anders, J., Hubbard, B., Hussain, A. & Heeley, M. (2015). Research data sharing: developing a stakeholder-driven model for journal policies. *Journal of the Association for Information Science and Technology*, 66(12), 2445-2455. Disponível na WWW: <http://onlinelibrary.wiley.com/doi/10.1002/asi.23336/abstract>

Suber, P. (2008). *Open Access News*. Disponível na WWW: <http://legacy.earlham.edu/~peters/fos/2008/08/greengold-oa-and-gratislibre-oa.html>

Suber, P. (2009). Ten challenges for open-access journals. *SPARC Open Access Newsletter*, 138. Disponível na WWW: <http://legacy.earlham.edu/~peters/fos/newsletter/10-02-09.htm>

Suber, P. (2012). *Open Access*. Cambridge: MIT Press. Disponível na WWW: https://mitpress.mit.edu/sites/default/files/9780262517638_Open_Access_PDF_Version.pdf

Sunlight Foundation (2010). Ten Principles for Opening Up Government Information. Disponível na WWW: <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>

Sunlight Foundation (2013). Public Policy for Public Data. Disponível na WWW: <http://whatworkscities.sunlightfoundation.com/#section-h2-02>

Sunlight Foundation (2014). *Guidelines for Open Data Policies*. Disponível na WWW: <https://sunlightfoundation.com/opendataguidelines/>

Sunlight Foundation. *Open Data Policy Collection*. [Consultado a 12-05-2017]. Disponível na WWW: <http://www.opendatapolicies.org/browse/>

Swan, A. (2010). The open access citation advantage: studies and results to date. Disponível na WWW: http://eprints.soton.ac.uk/268516/2/Citation_advantage_paper.pdf

Swan, A. (2012). Diretrizes para as Políticas de desenvolvimento do acesso aberto. UNESCO: Paris. Disponível na WWW: <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>

Tauberer, Joshua (2014). Open Government Data: The Book (2ª ed.). Disponível na WWW: <https://opengovdata.io/2014/8-principles/>

Taunabaum, G. Implementing an Open Data Policy: a SPARC Primer for Research Funders. Disponível na WWW: <https://sparcopen.org/wp-content/uploads/2016/01/sparc-open-data-primer-final.pdf>

Transparência e Integridade Associação Cívica (2016). *Índice de Transparência Municipal*. Disponível na WWW: <https://transparencia.pt/indice-de-transparencia-municipal/>

The UN Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014a). *UN Data Revolution*. Disponível na WWW: <http://www.undatarevolution.org/>

The UN Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014b). A world that counts: mobilising the data revolution for sustainable development. Disponível na WWW: <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>

The World Bank (2016). *Open Government Data Toolkit*. Disponível na WWW: <http://opendatatoolkit.worldbank.org/en/index.html>

Ubaldi, B. (2013). Open Government Data: Towards empirical analysis of open government data initiatives. Disponível na WWW: http://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en

UK Data Archive (2017). *Create & Manage Data: Research Data Lifecycle*. Disponível na WWW: <http://www.data-archive.ac.uk/create-manage/life-cycle>

UK. The National Archives. Open Government Licence for Public Sector Information. Disponível na WWW: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

UNESCO (2015). Concepts of Openness and Open Access. *Open Access for Researchers*. Paris: UNESCO. Disponível na WWW: <http://unesdoc.unesco.org/images/0023/002322/232207E.pdf>

União Europeia (2013). *REGULAMENTO (UE) Nº 1291/2013 DO PARLAMENTO EUROPEU E DO CONSELHO de 11 de dezembro de 2013*. Jornal Oficial da União Europeia. Disponível na WWW: http://ec.europa.eu/research/participants/data/ref/h2020/legal_basis/fp/h2020-eu-establact_pt.pdf

União Europeia (2017). *European Union Open Data Portal*. [Consultado a 12-05-2017]. Disponível na WWW: <http://data.europa.eu/euodp/en/data>

United Nations Statistics Division (2017). *UN data: a world of information*. [Consultado a 13-05-2017]. Disponível na WWW: <http://data.un.org/Search.aspx?q=Portugal&t=Data>

United States General Services Administration. *Data.gov*. Disponível na WWW: <https://www.data.gov/open-gov/>

Universidade do Minho (2017). *Open Access*. Disponível na WWW: https://openaccess.sdum.uminho.pt/?page_id=352

Universidade do Porto. Faculdade de Engenharia (2017). *Dendro: The convergence between data management and description*. Disponível na WWW: <http://dendro.fe.up.pt/blog/index.php/dendro/>

Welcome Trust (2013). *Open access policy*. Disponível na WWW: <https://wellcome.ac.uk/funding/managing-grant/open-access-policy>

Whyte, A. & Donnelly, M. (2016). Common directions in research data policy: a briefing for institutions. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Disponível na WWW: <http://www.dcc.ac.uk/resources/briefing-papers/common-directions-research-data-policy/common-directions-research-data-pol>

World Council on City Data (2016). *Data Portal*. [Consultado a 12-04-2017]. Disponível na WWW: <http://open.dataforcities.org/>

World Wide Web Consortium [W3C], (2014). *Data Catalog Vocabulary (DCAT)*. Disponível na WWW: <https://www.w3.org/TR/vocab-dcat/>

World Wide Web Consortium [W3C], (2015). *Linked Data*. Disponível na WWW: <https://www.w3.org/standards/semanticweb/data>

World Wide Web Foundation (2014, maio). Towards common methods for assessing open data: workshop report & draft framework. Apresentado em Towards common methods for assessing open data 2014, 8 - 9 Maio. Nova Iorque, EUA. Disponível na WWW: <http://opendataresearch.org/sites/default/files/posts/Common%20Assessment%20Workshop%20Report.pdf>

World Wide Web Foundation (2017). *ODB Global Report Fourth Edition*. Disponível na WWW: <http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf>

Zuiderwijk, A., Gasco, M., Parycek, P. & Janssen, M. (2014). Special issue on transparency and open data policies: guest editors' introduction. *Journal of theoretical and applied electronic commerce research*, 9(3) I-IX. Disponível na WWW: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-18762014000300001

Zuiderwijk, A. & Janssen, M. (2014). Open data policies, their implementation and impact: a comparison Framework. *Government Information Quarterly*, 31(1), 17-29. Disponível na WWW: <http://www.sciencedirect.com/science/article/pii/S0740624X13001202>

LISTA DE SIGLAS

AA - Acesso Aberto
AMA - Agência para a Modernização Administrativa
APC - Article Processing Charge
API - Application Programming Interface
APIS - Arquivo Português de Informação Social
BOAI - Budapest Open Access Initiative
CA - Ciência Aberta
CADA - Comissão de Acesso aos Documentos Administrativos
CC - Creative Commons
CERN - European Organization for Nuclear Research
CML - Câmara Municipal de Lisboa
DCAT - Data Catalog Vocabulary
DCC - Digital Curation Center
DDI - Data Documentation Initiative
DGA - Dados Governamentais Abertos
DMP - Data Management Plan
DOAJ - Directory of Open Access Journals
DOI - Digital Object Identifier
EAGDA - Expert Advisory Group on Data Access
e-IRG - e-Infrastructure Reflection Group
EOSC - European Open Science Cloud
ERAC - European Research Area and Innovation Committee
EU - União Europeia
EUA - Estados Unidos da América
FCCN - Fundação para a Computação Científica Nacional
FCT - Fundação para a Ciência e Tecnologia
FI - Fator de Impacto
FOI - Freedom of Information
FOSTER - Facilitate Open Science Training for European Research
GBIF - Global Biodiversity Information Facility
ICSU - International Council for Science
INE - Instituto Nacional de Estatística
IPCB - Instituto Politécnico de Castelo Branco
ITM - Índice de Transparência Municipal
JISC - Joint Information Systems Committee
MCTES - Ministério da Ciência, Tecnologia e Ensino Superior
NIH - National Institutes of Health
OAI-PMH - Open Archive Initiative – Protocol for Metadata Harvesting
OD4D - Open Data for Development

ODC - Open Data Commons
ODIN - Open Data Inventory
ODRA - Open Data Readiness Assessment
OECD - Organização para a Cooperação e Desenvolvimento Económico
OGD - Open Government data
OGL - Open Government Licence
OGP - Open Government Partnership
OGWG - Open Government Working Group
OKI - Open Knowledge International
OpenDOAR - Directory of Open Access Repositories
PLOS - Public Library of Science
PSI - Public Sector Information
RCAAP - Repositório Científico de Acesso Aberto de Portugal
RDA - Research Data Alliance
RDF - Resource Description Framework
Re3data - Registry of Research Data Repositories
ROARMAP - Registry of Open Access Repository Mandates and Policies
SARDC - Serviço de Alojamento de Repositório de Dados Científicos
SCOAP3 - Sponsoring Consortium for Open Access Publishing in Particle Physics
SNIAmb - Sistema Nacional de Informação de Ambiente
SOAP - Study of Open Access Publishing
SPARC - Scholarly Publishing and Academic Resources Coalition
W3C - World Wide Web Consortium

ÍNDICE DE FIGURAS

Fig. 1. Os pilares da Ciência Aberta.	17
Fig. 2. As 5 escolas de pensamento das mudanças operadas pela Ciência Aberta.....	18
Fig. 3. Características e indicadores do Open Science Monitor.	20
Fig. 4. Taxonomia da Ciência Aberta.	22
Fig. 5. Paradigmas da investigação científica.	25
Fig. 6. Cadeia de abastecimento dos Dados Massivos.	27
Fig. 7. A cauda longa dos Dados Massivos.	28
Fig. 8. Ciclo de vida dos dados de investigação.....	31
Fig. 9. Tipos de dados científicos.....	34
Fig. 10. Esquema de classificação das 5 estrelas para Dados Abertos Ligados.....	35
Fig. 11. Pilares dos Dados Governamentais Abertos.	40
Fig. 12. Perspetivas da abertura dos dados governamentais.	44
Fig. 13. Esquema de classificação das 5 estrelas para as iniciativas de abertura de dados que envolvem os utilizadores.....	45
Fig. 14. Esquema de classificação das 5 estrelas para os portais de dados.	50
Fig. 15. Princípios FAIR Data.....	61
Fig. 16. Estrutura de avaliação de políticas de Dados Abertos.	67

ÍNDICE DE QUADROS

Quadro 1. Princípios de abertura de dados de investigação mais recomendados.....	33
Quadro 2. Princípios de abertura mais recomendados pelas iniciativas de Dados Governamentais Abertos.	43
Quadro 3. Visão geral sobre o contexto, conteúdo e desempenho da política de Dados Abertos em Portugal com base no esquema delineado por Zuiderwijk e Janssen (2014).	85
Quadro 4. Comparação dos resultados de abertura de dados atribuídos pelas várias iniciativas de avaliação.....	95
Quadro 5. Portais de dados abertos existentes em Portugal.	97

ÍNDICE DE GRÁFICOS

Gráfico 1. Número de <i>datasets</i> publicados no SARDC nos anos seguintes à sua criação.....	80
Gráfico 2. Maiores fornecedores de dados do portal nacional.	87
Gráfico 3. Tipologia dos fornecedores de dados do portal nacional.	88
Gráfico 4. Tipologia das categorias de dados do portal nacional.	88