

Os desafios dos dados de pesquisas abertos

Open research data challenges

Desafíos de los datos de investigaciones abiertos

Pascal Aventurier | pascal.aventurier@avignon.inra.fr

Institut National de la Recherche Scientifique (France). Centre de Recherche PACA. Avignon, France.

Maria de Cléofas Faggion Alencar | cleofas.alencar@embrapa.br

Empresa Brasileira de Pesquisa Agropecuária. Embrapa Meio Ambiente. Jaguariúna, SP, Brasil.

Resumo

A gestão, a disponibilização e a preservação dos dados de pesquisas são preocupações cada vez mais centrais para os pesquisadores, as instituições de pesquisas e também para as agências de fomento e revistas científicas. Todos esses têm o objetivo de desenvolver boas práticas de pesquisa que permitam a verificação e a reutilização dos dados, a valorização dos trabalhos de construção de conjuntos de dados e o desenvolvimento de novas maneiras de pesquisar utilizando dados já produzidos. Este artigo apresenta as diferentes etapas do ciclo de vida dos dados e mostra as iniciativas internacionais e nacionais tomadas para desenvolver esse campo que tem grande interesse para a área da saúde.

Palavras-chave: Dados de pesquisas; Repositórios de dados; *Data journals*; Dados abertos; Ciência aberta.

Abstract

Increasingly, the management, availability and preservation of research data concern researchers, research institutions as well as research funding agencies and scientific journals. All of them aim to develop good research practices allowing the data confirmation and their reuse, the improvement of the dataset building and the development of new ways of searching through the use of data already produced. This paper presents the different stages of the data life cycle and shows the international and national initiatives to develop these important issues to the health field.

Keywords: Research data; Data repositories; *Data journals*; Open data; Open science.

Resumen

La gestión, la disponibilidad y conservación de datos procedentes de la investigación constituyen preocupaciones cada vez mayores para los investigadores, instituciones de investigación, organismos de financiación y revistas científicas. Todos tienen el objetivo de desarrollar buenas prácticas de investigación, permitiendo la verificación y la reutilización de los datos, la valorización de los trabajos de construcción del conjunto de datos y el desarrollo de nuevas maneras de hacer investigaciones utilizando datos ya producidos. Este artículo presenta las diferentes etapas del ciclo de vida de los datos y muestra las iniciativas internacionales y nacionales para el desarrollo de este campo que tiene gran utilidad para la área de la salud.

Palabras clave: Datos de investigación; Repositorios de datos; *Data journals*; Datos abiertos; Ciencia abierta.

INFORMAÇÕES DO ARTIGO

Contribuição dos autores:

Concepção e desenho do estudo: Pascal Aventurier
Aquisição, análise ou interpretação dos dados: Pascal Aventurier
Redação do manuscrito: Pascal Aventurier, Maria de Cléofas Faggion Alencar
Revisão crítica do conteúdo intelectual: Maria de Cléofas Faggion Alencar

Declaração de conflito de interesses: não há nenhum conflito de interesse.

Fontes de financiamentos: não houve.

Agradecimento/Contribuições adicionais: os autores agradecem a Juliana Reis, que nos forneceu aconselhamento editorial.

Histórico do artigo: Submetido: 08.jan.2016 | Aceito: 25.jan.2016 | Publicado: 30.set.2016

Apresentação anterior: não houve.

Licença CC BY-NC atribuição não comercial. Com essa licença é permitido acessar, baixar (download), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos, desde que para uso não comercial e com a citação da fonte, conferindo os devidos créditos de autoria e menção à Reciiis. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores.

Introdução

A maneira de conduzir a ciência está evoluindo, puxada pelo movimento *Open Science* ou Ciência Aberta, que inclui um conjunto de práticas sobre a utilização da internet e das ferramentas da web social para mudar de maneira global a abordagem acadêmica, desde a formulação das questões e das hipóteses até a difusão dos resultados de pesquisas. Outros princípios da *Open Science* são apresentados na Figura 1. Para cada princípio existem ferramentas que os pesquisadores estão progressivamente adotando como, por exemplo, os sites para *Open Peer Review*, repositórios abertos e institucionais de documentos ou mesmo ferramentas abertas relacionadas ao ensino.

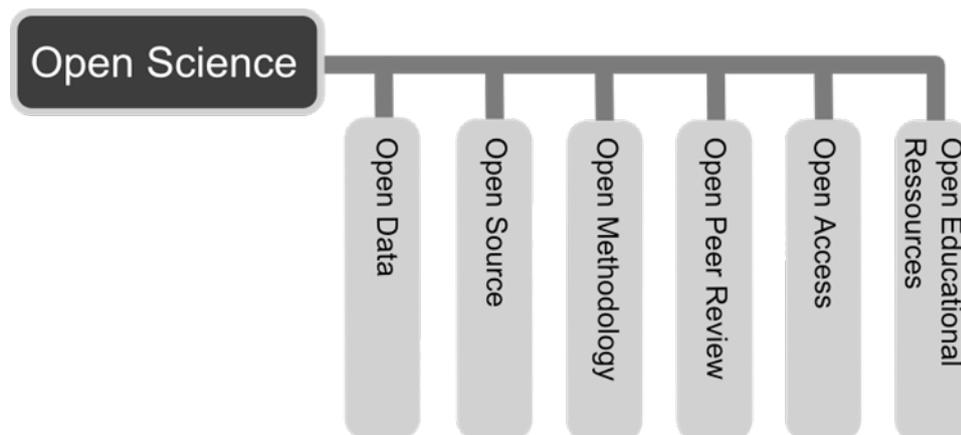


Figura 1 - Os componentes da *Open Science*
Fonte: Wikipediaⁱⁱ

O campo dos dados de pesquisas evoluiu em três grupos: as preocupações com a disponibilização dos dados e sua permanência, a exploração dos dados e os dados ligados a uma publicação como elementos de prova. Em 1957, houve uma iniciativa internacional do International Council of Science (ICSU), que criou o World Data Centre (WDC)ⁱⁱⁱ para coletar e distribuir dados. Em 1962, John W. Tukey publicou um artigo intitulado *The future of data analysis*¹ que dá início ao papel do “*data scientist*”. De 2003 a 2004, as agências de fomento à pesquisa como o National Institute of Health (Estados Unidos) e a The Royal Society (Reino Unido) começaram a estabelecer políticas para disponibilizar esses dados de pesquisas. A Declaração de Berlim sobre o Acesso Livre ao Conhecimento nas Ciências e Humanidades de 2003^{iv} definiu acesso livre e também acesso aos dados de pesquisas. A primeira política de editores sobre dados de pesquisas começou com o International Committee of Medical Journal Editors (ICMJE), que solicitou a disponibilização de ensaios sobre medicamentos^v.

Compartilhar os dados de pesquisas é uma necessidade. Hoje, a sociedade enfrenta desafios científicos muito mais complexos que pressupõem a interdisciplinaridade, colaborações entre laboratórios de pesquisa, permuta de informações e de competências. Com isso, novas competências associadas ao trabalho com os dados de pesquisas estão surgindo para corroborar com o descobrimento de novos conhecimentos a partir desses volumes de dados.

i Repositório Arca, da Fiocruz, disponível em: <<http://www.arca.fiocruz.br/>> e Pubmed Central em Ciência da Biologia, disponível em: <<http://www.ncbi.nlm.nih.gov/>>.

ii Disponível em: <https://en.wikipedia.org/wiki/Open_science>.

iii Disponível em: <<https://www.icsu-wds.org/>>.

iv Disponível em: <https://openaccess.mpg.de/67693/BerlinDeclaration_pt.pdf>.

v Disponível em: <<http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>>.

Outros desafios da disponibilização de dados são a gestão e a exploração do volume de dados produzidos. Um dos estudos sobre o tema mostra que o volume do *data universe* duplica a cada dois anos e que, em 2020, os dados de pesquisas atingirão 44 zettabytes, ou seja, 44 trilhões de gigabytes. Desde 1982, o Genbank, banco de dados das sequências de DNA, se duplica aproximadamente a cada 18 meses^{vi}. Em sua versão de abril 2016, contém mais de 211 bilhões de bases de nucleotídeos e mais de 193 milhões de sequências diferentes (Figura 2).

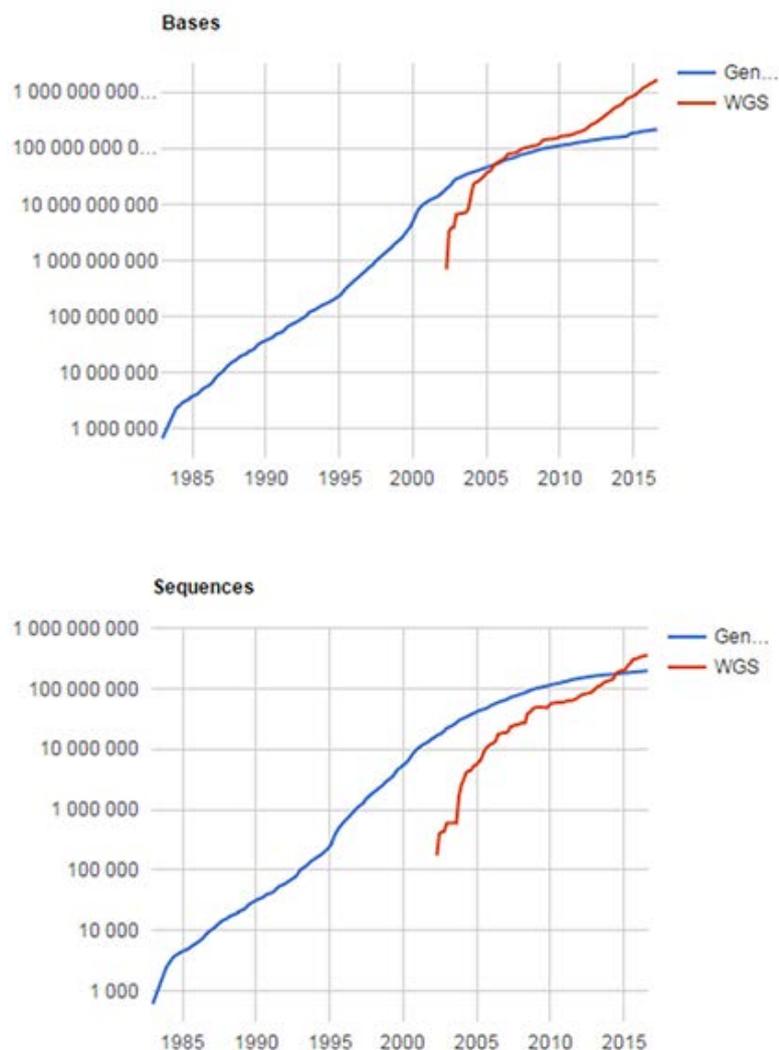


Figura 2 - O crescimento do Genbank
Fonte: Genbank (2016).

No contexto da ciência aberta e do aumento do volume de dados, os dados de pesquisas abertos apresentam potencial estratégico em diferentes níveis de uma organização para se explorar e valorizar seus conteúdos, além de acompanhar a evolução da atitude filantrópica com o compartilhamento público do conhecimento.

As instituições é que definem as condições de abertura de seus dados (licença, convenções de uso, citações), primeiramente em seus próprios laboratórios e, em seguida, alinhando-se às novas políticas das fundações de financiamento à pesquisa internacionais, possibilitando mais chances dos projetos de pesquisa serem aprovados. Na Europa e nos Estados Unidos, por exemplo, vários órgãos governamentais de subvenções exigem que os projetos disponibilizem os resultados abertamente.

vi Disponível em: <<http://www.ncbi.nlm.nih.gov/genbank/statistics>>.

Da mesma forma, os laboratórios poderiam ser mais valorizados, publicando seus conjuntos de dados para que outros possam utilizá-los e assim alcançar colaborações ou reconhecimento. Um exemplo importante desse fato são as equipes de engenheiros que não publicam artigos, mas produzem dados para a pesquisa e dessa forma estão sendo mais valorizados.

Neste artigo, pretende-se abordar as questões de gestão através do ciclo de vida dos dados de pesquisas, uma ferramenta muito útil com a qual o plano de gestão de dados define as condições de produção, de conservação, de reutilização dos dados no próprio projeto de pesquisa. A citação dos dados também é importante e necessária; por isso, ao disponibilizar os dados em um repositório, sugere-se definir a sua forma e atribuir um DOI - *Digital Object Identifier*^{vii}.

Os dados de pesquisas estão intrinsecamente ligados aos artigos e periódicos científicos. Hoje, os editores exigem a disponibilização dos dados para avaliarem as publicações e reproduzirem os resultados. Aliás, a premissa de validação de uma pesquisa em sua forma de publicação é exatamente a possibilidade do leitor conseguir reproduzi-la.

Desde 2010, estão disponíveis os *data journals*, periódicos que publicam *data papers*, um tipo de artigo que descreve um conjunto de dados, as condições de produção e o interesse científico. Alguns exemplos de *data journals* são: [Geoscience Data Journal](#) (2012), [Scientific Data](#) (2013) e [Journal of Open Archaeology Data](#) (2011)².

Para os pesquisadores, a publicação dos dados pode valorizar ainda mais os artigos da pesquisa, no momento em que outros pesquisadores os acessarem e os citarem.

A proposta aqui é também fazer um mapa da situação atual da disponibilização dos dados de pesquisas que, em pouco tempo, têm evoluído muito, promovendo sua publicação e auxiliando a validação das pesquisas.

Definições

O Relatório da OECD^{viii}, sigla em inglês da Organização para a Cooperação e Desenvolvimento Econômico, descreve a expressão “dados de pesquisas” como “registros factuais usados como fonte primária para a pesquisa científica e que são comumente aceitos pelos pesquisadores como necessários para validar os resultados do trabalho científico”. Entretanto, o que se observa é que a amplitude do que se entende por dados de pesquisas sugere um conceito complexo que pode se manifestar numa multiplicidade de formas.

Seguindo a categorização proposta pelo National Science Board^{ix}, os dados podem ser discriminados pela sua natureza ou origem³:

- Observacionais, obtidos por meio de observação;
- Computacionais, resultantes de execução de modelos computacionais ou de simulação;
- Experimentais, provenientes de situações controladas em bancadas de laboratórios.

A Universidade de Melbourne⁴ definiu dados como fatos, observações ou experiências baseadas em argumentos, teorias ou testes. Os dados podem ser digitais, descritivos ou visuais. Também, podem ser brutos, analisados, experimentais ou observacionais. Eles incluem cadernos de laboratórios ou relatórios de experiência; dados primários coletados, questionários; vídeos ou áudios, modelo computacional, imagens e respostas de levantamentos. As informações sobre a produção dos dados devem ser também disponibilizadas, isto é, devem ser explicitados o local, a data, o protocolo, a ferramenta para a análise e o código computacional.

vii Digital Object Identifier (DOI)³⁰ “é um padrão para identificação de documentos em redes de computadores, como a Internet”.

viii Disponível em: <<http://www.oecd.org/>>.

ix Disponível em: <<http://www.nsf.gov/nsb/>>.

Existem várias outras definições de dados de pesquisas em função do estado de processamento e do material estudado. O *Australian National Data Service*⁵ menciona que os dados brutos e sem processamento de um fenômeno em particular devem ser processados para a correção do cálculo da calibração estatística; outros são derivados de um ponto de vista específico do conjunto de dados brutos. Além disso, existem os dados textuais, como publicações ou outros documentos.

O livro *The fourth paradigm*⁶ expande a visão do pioneiro cientista da computação Jim Gray⁷, mostrando que os dados de pesquisas são produzidos por diferentes meios automáticos como a programação, a coleta por instrumentos etc. O conceito de *e-science* promove o conjunto de metadados e as ferramentas não só para analisar dados e realizar a mineração de textos, como visualizar e disseminar dados. O modo de funcionamento da pesquisa é fundamentalmente diferente. Ele é conduzido pelos dados: os resultados são obtidos a partir do volume de dados sem hipótese inicialmente e, depois, são utilizados para elaborar e validar hipóteses.

As iniciativas institucionais e temáticas para manutenção dos dados de pesquisas em vários níveis são muito importantes porque definem os contextos de produção e de valorização desses dados. São guias e incentivadoras para os pesquisadores.

Agências de fomento

As principais agências de fomento de pesquisa exigem a disponibilização dos dados em repositórios.

Nicol et al.⁸, em relatório de 2013 intitulado *Open Data Access Policies and Strategies in the European Research Area and Beyond*, analisaram estratégias para incentivar as instituições e agências de fomento na União Europeia e também no Brasil, Canadá, Japão e nos Estados Unidos para viabilizar o acesso aberto dos dados de pesquisas.

O relatório do OECD de 2007, *Principles and Guidelines for Access to Research Data from Public Funding*⁹, indica que os dados de pesquisas financiadas com recursos públicos deveriam estar facilmente — e rapidamente — disponíveis na internet porque os custos marginais são muito baixos em comparação com o custo de sua produção.

O Digital Curation Center (DCC) estudou, em 2014, as políticas de gestão de dados das agências de fomento dos RCUK (Research Councils UK)¹⁰ sob o ponto de vista da cobertura, condições e apoio fornecido (Figura 3). Valen, D. e K. Blanchat¹¹ também estudaram o conteúdo dos órgãos de financiamento dos Estados Unidos pesquisados dirigidos pela OSTP (US Office of Science and Technology Policy).

● Full Coverage ● Partial Coverage ○ No Coverage

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	●	○	●	○	○	○
BBSRC	●	●	●	●	●	●	●	●	●	○	●
CRUK	●	●	●	●	●	●	●	○	●	○	○
EPSRC	●	●	●	○	●	●	●	○	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	○
MRC	●	●	●	●	●	●	○	○	●	○	○
NERC	●	●	●	●	●	●	●	●	●	●	○
STFC	●	●	●	●	●	●	●	○	●	○	○
Wellcome Trust	●	●	●	●	●	●	●	●	●	○	●

Figura 3 - Políticas de gestão de dados dos Conselhos de Pesquisa do Reino Unido analisadas pela Fundação de financiamento do Reino Unido (Digital Curation Center)

Fonte: Digital Curation Center (2016).

Nesses dois casos, os eixos exigidos são:

- O compartilhamento de dados;
- Um plano de gestão de dados;
- Um período de compartilhamento (permanência de manutenção dos dados disponíveis).

Dietrich et al.¹² mostram que, em 2012, houve uma diferença entre os objetivos de gestão e as realidades de implementação, em que muitas agências americanas de fomento à pesquisa têm exigências imprecisas e sem detalhes de implementação. Os autores mostram também que as oportunidades para os profissionais das bibliotecas aumentaram muito no sentido de apoiar os pesquisadores na gestão de dados.

Iniciativas institucionais e acionais

Muitas universidades americanas (MIT, Universidade da Califórnia), inglesas (Oxford, Edimburg) e australianas (Monash, Sidney) estão divulgando suas políticas para a gestão de dados de pesquisas e organizando apoio para o estabelecimento de um plano de gestão e curadoria de dados delegado às bibliotecas. Essas políticas universitárias apoiam as políticas das fundações de financiamento à pesquisa e o movimento de acesso aberto. Além disso, estão desenvolvendo ferramentas para estabelecer planos de gestão de dados como DMP Tools^x ou DMP Online^{xi}.

O Reino Unido criou um Centro de Curadoria Digital (DCC ou Centro para a Curação Digital^{xii}), e a Austrália o ANDS Agência Nacional para Serviços de Dados^{xiii}.

Iniciativas por temáticas

Alguns temas como a astronomia, a biologia e a agricultura já têm iniciativas que incentivam diferentes organizações a publicar dados. Entre elas, encontram-se:

- *The Bouchout Declaration for Open Biodiversity*^{xiv}: declaração para que os cientistas e as organizações declarem interesse em compartilhar dados sobre a biodiversidade. Em 25 maio de 2015, 92 organizações, incluindo três brasileiras e 202 indivíduos (entre eles 12 brasileiros) assinaram essa declaração;
- *Biosharing*^{xv}: catálogo de conjuntos de dados de pesquisas descritos para os *BioDBcore guidelines*^{xvi}, com normas utilizadas e parcialmente processadas com o apoio de Oxford University Press;
- *Research Data Alliance*^{xvii}: criada em março de 2013 pela Comissão Europeia, pelos Estados Unidos (NSF) e pela Austrália (ANDS), de modo a facilitar o compartilhamento dos dados científicos;
- *GODAN, Global Open Data for Agriculture and Nutrition*^{xviii}: criada em março de 2013, é uma

x Disponível em: <<https://dmp.cdlib.org/>>.

xi Disponível em: <<https://dmponline.dcc.ac.uk/>>.

xii Disponível em: <<http://www.dcc.ac.uk>>.

xiii Disponível em: <<http://ands.org.au/>>.

xiv Disponível em: <<http://www.bouchoutdeclaration.org/>>.

xv Disponível em: <<https://www.biosharing.org/biodbcore/>>.

xvi Disponível em: <<https://biosharing.org/databases/>>.

xvii Disponível em: <<https://rd-alliance.org/>>.

xviii Disponível em: <<http://www.godan.info/>>.

iniciativa para desenvolver colaborações e políticas para abertura de dados relacionados à agronomia e à nutrição.

- *The Hague Declaration on Knowledge Discovery in the Digital Age*^{xix}: assinada por mais de 160 organizações e 330 pesquisadores ou afiliados, essa declaração pretende não só promover melhor acesso aos fatos, dados e ideias para novos descobrimentos do conhecimento na era digital, como também ampliar o acesso e a análise de dados produzidos pela sociedade para resolver desafios como a mudança climática, a diminuição da biodiversidade e a globalização.

No Brasil

O Brasil não tem até agora iniciativas institucionais ou governamentais para incentivar, compartilhar, disponibilizar ou produzir os dados de pesquisas. Mesmo assim, os pesquisadores brasileiros estão seguindo as regras adotadas pelos editores internacionais quando seus artigos são aprovados, isto é, depositam os dados de pesquisas conforme as instruções para disponibilizá-los.

O Programa SciELO - *Scientific Electronic Library Online*, modelo de publicação científica, possui critérios, políticas e procedimentos de avaliação de periódicos científicos para a admissão e permanência de títulos na Coleção SciELO Brasil. É utilizado também como documento de referência para a avaliação de periódicos nas demais coleções nacionais e temáticas da Rede SciELO. Em documento de 2014¹³, indica política de disponibilização de dados de pesquisa no parágrafo 5.2.15, mencionando a seguir que passou a vigorar em 2015.

“A disponibilização dos dados das pesquisas utilizados nos artigos em repositórios de acesso aberto, seguindo padrões de registro que assegurem a autoria, o uso e citação dos dados assim como do artigo correspondente é recomendável, pois contribui para a replicabilidade das pesquisas e aumenta a visibilidade e as citações das pesquisas e dos periódicos.”

O barômetro *“Open Data Research”* da Web Foundation^{xx}, de Tim Berners-Lee, classifica o Brasil na categoria de “emergente e avançado”, que agrega países com programas de dados abertos de instituições ou uma agenda de política de acesso aberto considerados inovadores. Nessa categoria, o Brasil está classificado em 4º lugar, atrás da Espanha, do Chile e da República Checa, mas na frente da Itália.

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), a Rede Cariniana^{xxi} e a Universidade de Brasília (UNB) propõem o planejamento de um projeto para construção de um repositório de dados de pesquisas usando a ferramenta Dataverse, da Universidade de Harvard, para todas as áreas do projeto Jaiba Solar da Faculdade de Tecnologia da Universidade de Brasília¹⁴. A plataforma Dataverse é uma arquitetura de software livre para a publicação, citação, análise e preservação de dados de projetos de pesquisa, que permite um gerenciamento de dados seguro. O Repositório de Dados Científicos (DVN)^{xxii}, da Universidade Federal do Rio Grande do Sul (UFRGS), também está em fase de implementação.

O Centro de Informações Nucleares (CIN) publicou um guia de gestão de dados de pesquisas¹⁵ que tem como objetivo apresentar as informações fundamentais e as melhores práticas para a gestão de dados.

Europa

Na Europa, o programa *Horizon 2020* especifica que os dados de pesquisas necessários para validar os resultados apresentados em publicações devem ser depositados concomitantemente aos artigos, de

xix Disponível em: <<http://thehaguedeclaration.com/>>.

xx Disponível em: <<http://www.webfoundation.org/>>.

xxi Disponível em: <<http://cariniana.ibict.br/>>.

xxii Disponível em: <<http://unbral.nuvem.ufrgs.br/dvn/>>.

preferência em um repositório de dados. Não há a menor obrigação de deixá-los em acesso aberto, embora sejam muito bem-vindos e façam parte do *Open Data Research Pilot*⁶.

Existem casos em que não é possível compartilhar publicamente os dados como, por exemplo, por causa de sua confidencialidade. Nesses casos, os dados e os metadados necessários para a avaliação dos trabalhos devem estar acessíveis aos avaliadores. O projeto *Openaire Plus* utiliza o repositório Zenodo^{xxiii}, desenvolvido pelo Conseil Européen pour la Recherche Nucléaire (Cern), além de regras para estabelecer um plano de gestão de dados, e os disponibiliza para serem utilizados por outros pesquisadores, ou seja, compartilha da ideia de *open data*. O projeto piloto inclui informações sobre ferramentas que avaliam os resultados. Outros conjuntos de dados poderiam também ter sido incluídos, mas o projeto piloto não pode ser aplicado a todos os tipos de dados, especialmente dados sensíveis como os de pacientes que não podem ser compartilhados ou apropriados. Alguns dos atores, como a organização Datacite ou o Codata (Committee on Data for Science and Technology)^{xxiv}, criaram uma infraestrutura global de reflexão sobre dados de pesquisas.

A organização Datacite^{xxv} ajuda os data center, os periódicos e os pesquisadores a desenvolverem métodos para identificar, citar dados e outros assuntos científicos. O principal serviço da Datacite consiste em fornecer identificação permanente como os números DOI para os objetos digitais e também para os conjuntos de dados. O serviço é reservado aos membros da Datacite. Como os artigos já possuem DOI, os DOIs para os conjuntos de dados permitiriam a citação e a ligação entre um artigo e um conjunto de dados.

Algumas organizações de bibliotecários e outros especialistas em dados estão muito envolvidos em diferentes ações com os dados científicos, como por exemplo a Force¹¹ ^{xxvi} e a Liber Europe^{xxvii}. O Projeto Foster da União Europeia^{xxviii} tem como objetivo o desenvolvimento de ações de formação em Open Science e o fornecimento de princípios, serviços e um grupo de trabalho sobre esse assunto.

A União Europeia também criou um grupo de especialistas da *Science Europe* para compartilhar políticas sobre coleta, qualidade, preservação e compartilhamento ou reutilização dos dados^{xxix}.

O movimento de dados abertos (Open data movement)

O movimento de dados abertos é radicalmente diferente do processo de dados de pesquisas abertos, mas poderá funcionar como alavanca para disponibilizar os dados de pesquisas e também levar dados governamentais disponíveis para a pesquisa científica. O movimento se propõe a incentivar os governos que querem mostrar transparência e disponibilização dos dados públicos.

O G8 incentivou a disponibilização dos dados em 2013 através do G8 Open Data Charter e reconheceu que algumas áreas científicas como as de segurança alimentar, desenvolvimento e observação da terra deveriam ter acesso aberto como uma opção preestabelecida¹⁷.

Os governos disponibilizam cada vez mais dados em seus portais governamentais. O portal Brasil tem cerca de 1.000 conjuntos de dados^{xxx}, enquanto os Estados Unidos possuem mais de 130.000.

xxiii Disponível em: <<https://zenodo.org/>>.

xxiv Disponível em: <<http://www.codata.org/>>.

xxv Disponível em: <<https://www.datacite.org/>>.

xxvi Disponível em: <<https://www.force11.org/>>.

xxvii Disponível em: <<http://libereurope.eu/>>.

xxviii Disponível em: <<https://www.fosteropenscience.eu/>>.

xxix Disponível em: <<http://www.scienceeurope.org/policy/working-groups/Research-Data>>.

xxx Disponível em: Brasil <<http://dados.gov.br/>>, US <<http://www.data.gov/>>.

As licenças de difusão de dados de pesquisas

Para favorecer a reutilização dos dados é necessário associar uma licença de difusão exigida pelos repositórios de dados ou pelo próprio periódico. No caso de disponibilizar os dados de pesquisas, os autores poderiam conservar os dados privados para valorizá-los em outras publicações. Embora a questão do embargo seja muito importante, deveria ser objeto de negociação entre os editores dos periódicos e os autores.

As principais licenças são do *Creative Commons 4.0*^{xxx1}, com seis licenças diferentes (chamadas CCs) de compartilhamento de obras e uma condição mínima de utilização livre com paternidade da obra (Licença CC-BY 4.0).

As outras opções são: permitir as adaptações do seu próprio trabalho (Sim/Não/Sim, desde que outros compartilhem igualmente) ou permitir usos comerciais do seu trabalho (Sim/Não).

As CCs são licenças em “três camadas”: o texto legal que os advogados conhecem, o resumo explicativo que sumariza e explica alguns dos termos e uma versão digital da licença que permite aos motores de busca utilizá-la.

Na CCO de “dedicação ao domínio público”, a pessoa que associa um trabalho ao resumo dessa licença dedica o trabalho ao domínio público, renunciando a todos os seus direitos sob as leis de direito de autor e/ou de direitos conexos referentes ao trabalho, em todo o mundo, na medida permitida por lei.

A pessoa pode copiar, modificar, distribuir e executar o trabalho, mesmo para fins comerciais, tudo sem pedir autorização. Essa licença é cada vez mais utilizada pelos repositórios como, por exemplo, o Dryad Digital Repository e os periódicos *BioMed Central e Nature Publishing Group*. As licenças ODC-By^{xxxii} (*Open Database Commons*) autorizam a cópia, a utilização, a redistribuição e a modificação dos dados, mas obrigam a indicação do nome do autor do conjunto de dados de origem. A licença ODC-By é próxima à licença CC-BY e tem sido utilizada por muitos editores científicos.

A licença ODC-ODBL^{xxxiii} impõe a redistribuição gratuita e se aproxima mais da licença CC-BY-AS.

A PDDL (*Public domain dedication and license*)^{xxxiv} é uma licença muito parecida com a licença CCO do tipo domínio público, em que os autores abandonam todos os direitos morais.

Onde publicar os dados de pesquisas?

Dependendo da área de investigação, os dados de pesquisas podem frequentemente ser depositados em um ou mais repositórios que permitem acesso aos dados. Esses repositórios podem ter requisitos específicos como:

- Assunto domínio/pesquisa;
- Licença para acesso aos dados e seu reuso;
- Formato de arquivo e estrutura de dados;
- Metadados de descrição;
- Embargo;
- Preços;
- Ligação com periódicos;
- Atribuição de identificação como o DOI.

xxx1 Disponível em: <https://creativecommons.org/licenses/?lang=pt_BR>.

xxxii Disponível em: <<http://opendatacommons.org/licenses/by/>>.

xxxiii Disponível em: <<http://opendatacommons.org/licenses/odbl/>>.

xxxiv Disponível em: <<http://opendatacommons.org/licenses/pddl/>>.

Essas ferramentas são novas e as suas funcionalidades estão evoluindo.

Além das indicações de repositórios de dados de pesquisas dos próprios editores de periódicos, as bibliotecas universitárias mantêm listas de repositórios como faz, por exemplo, a Universidade de Oregon^{xxxv}.

Um repositório de repositórios, desenvolvido pela iniciativa Datacite, Res3data.org (Figura 4), identificou 1.500 repositórios de dados de pesquisas em abril de 2016 que podem ser acessados pela navegação no site ou por programação^{xxxvi}. O Res3data.org não permite acesso direto aos conjuntos de dados, mas somente aos repositórios.



Search for Repositories (1223 Reviewed Repositories)

Figura 4 - Registro de repositórios de dados de pesquisas
Fonte: Datacite, 2016.

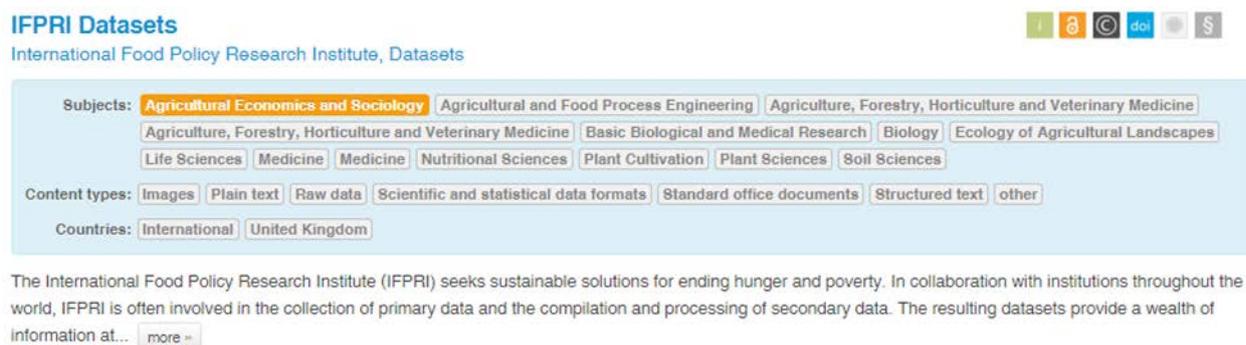


Figura 5 - Interface do Re3data.org e descrição de um conjunto de dados
Fonte: Datacite, 2016.

No entanto, uma de suas interfaces possibilita a busca por assunto (Figura 5). O serviço Datacite permite ainda acessar diretamente os conjuntos de dados, como mostra a Figura 6, através do site search.datacite.org^{xxxvii}. O serviço é aberto, mas contém somente os conjuntos de dados de repositórios registrados na organização Datacite.

xxxv Disponível em: <<http://library.uoregon.edu/datamanagement/repositories.html>>.

xxxvi Disponível em: <<http://www.re3data.org/2015/03/merger-and-first-version-of-api/>>.

xxxvii Disponível em: <<http://search.datacite.org/ui?>>.

The screenshot shows the DataCite Metadata Search beta interface. At the top, there is a search bar with the word 'dengue' entered and a 'Search' button. To the right of the search bar are links for 'Options', 'Advanced Search', 'About Us', 'Contact', and 'Help'. On the left side, there is a 'Filter' sidebar with various categories like 'allocator', 'datacentre', 'prefix', 'resourceType', 'contributor', 'creator', 'publicationYear', 'publisher', and 'language'. The main content area displays search results for 'dengue'. It shows '221 documents found in 46ms' and 'Page 1 of 23'. The first result is titled 'Behavioral responses of Aedes aegypti to insecticides using two assays systems and the influence of insecticide resistance mechanisms /' with a DOI of 10.14457/KU.THE.2008.147 Dataset and subject 'Dengue'. The second result is 'Effects of compounds from CLINACANTHUS NUTANS and ANDROGRAPHIS PANICULATA on dengue virus 2 infection and prostaglandin e2 # 2' with a DOI of 10.14457/KU.THE.2011.10 Dataset and subject 'Dengue virus'. The third result is 'Influenza : grippe. - Fièvre dengue # 3' with a DOI of 10.5169/SEALS-191348 and subject 'Dengue'. The fourth result is 'Dengue ICU data # 4' with a DOI of 10.6084/M9.FIGSHARE.1360094 Dataset and subject 'Dengue ICU data'.

Figura 6 - Exemplo de busca na Datacite com a palavra dengue
 Fonte: Datacite, 2016.

Existem alguns critérios para se escolher um repositório conforme as ações relacionadas. Para dados depositados em conjunto com um artigo de periódico, pode-se escolher um:

- Repositório multidisciplinar indicado pelo periódico;
- Repositório exigido pela fundação de financiamento à pesquisa, repositório da sua instituição, explicando aos editores do periódico a escolha ou exigência da instituição afiliada;
- Repositório da sua temática explicando aos editores do periódico a escolha;
- Repositório de menor custo.
- Se o conjunto de dados não estiver associado a uma publicação, a melhor prática é depositá-lo em um repositório e publicar um *data paper* descrevendo os dados em um periódico específico para isso denominado *data journal*.
- Nessa situação, os critérios para a escolha são:
- Repositório da sua própria instituição;
- Repositório baseado na temática ou um repositório multidisciplinar;
- Repositório com serviços de curadoria de informação, atribuição de DOI etc.;
- Repositório de menor custo.

O Quadro 1 a seguir apresenta cinco repositórios de dados de pesquisas com a descrição de suas características. Outras ferramentas, como Github^{xxxviii}, por exemplo, são mais adequadas para depositar o código computacional.

xxxviii Disponível em: <<https://github.com/>>.

Quadro 1 - Características de alguns repositórios

Nome	Tipo	Temática	Número de depósitos	Editores	Formato	Licença
DRYAD	-	biociências	2.521 conjuntos de dados, 6.888 data files	Dados estão associados com os artigos de 182 periódicos. Alguns editores: American Society of Naturalist, BioMed Central, Ecological Society of America	Hora de cálculo, fichas planos, imagens, mapas	Creative Commons Zero (CC0)
FIGSHARE	Privada (Digital Sciences) parceria com universidades	Todas	-	-	Imagens, figuras, hora de cálculo, conjunto de dados, mídia, conjunto de arquivos, artigos, posters, apresentações	Creative commons
PANGEA	Universitária/ União Europeia	Observação da Terra, dos oceanos e da atmosfera	Dados de mais de 100 projetos	Arquivos de periódicos: Earth System Science Data (ESSD)	Todos os tipos convertidos automaticamente	Creative commons
ICPSR	Inter-universidades	Ciências sociais	Nd	Nd	Arquivos de dados quantitativos, SPSS, SAS ou Stata. Arquivos de dados quantitativos. Formato texto ou Rich-Text Format (RTF). Arquivos de vídeo e áudio MPEG4, JPEG2000, .aif ou .wav	Específico para acesso livre

*Quadro 1b - Características de alguns repositórios - Conclusão
Comparação entre cinco repositórios de dados de pesquisa associados a publicações.*

Nome	Tipo	Temática	Número de depósitos	Editores	Formato	Licença
ZENODO	União Europeia CERN, EU FP7 projeto OpenAIREplus	Todas	Nd	Nd	Conjunto de dados e todos tipo de documentos de pesquisa. Gratuito até 2 Gb por depósito.	Licenças flexíveis baseadas no Creative Commons

Fonte: Elaborado pelo autor Pascal Avenirier (2013). Dados de pesquisa ligados a publicações: uma comparação de 4 principais ferramentas. Disponível em: <<https://publicient.hypotheses.org/425>>.

Planos de gestão de dados

Os planos de gestão de dados estão ligados às políticas voltadas para ela. A abertura dos dados não é um acesso inicial, mas o resultado de uma boa política de gestão dos dados como condição prévia. Esses planos são desenvolvidos pelas universidades, agências de fomento à pesquisa ou centro nacional de dados como o DCC ou o ANDS.

Os principais financiadores das pesquisas querem cada vez mais a abertura dos dados de pesquisas, de modo que, para se obter um financiamento, são exigidos os planos de gestão dos dados com suas metas e os planos das instituições de pesquisa.

Um Plano de Gestão de Dados (PGD) ou, em inglês, Data Magement Plan (DMP) é um documento oficial sobre a gestão dos dados desde o início do desenvolvimento da pesquisa até o final. O PGD descreve o tipo de dados, os métodos de coleta, os metadados e a conservação dos dados, com objetivo de facilitar a sua compreensão e a reutilização futura.

Os PGD incluem, de modo geral, os seguintes itens:

- Descrição do projeto;
- Autores e afiliações;
- Exigências das agências de fomento ou de uma instituição;
- Tipo de dados produzidos;
- Normas de metadados e dados;
- Regras para distribuição e reutilização;
- Planos para conservação e arquivamento.

Existem variações entre conteúdo e orientação dos PGD. Por exemplo, o *Digital Curation Center Checklist for a data management plan*¹⁸ tem foco nas exigências de agência de fomento porque associa o número do contrato, e não as regras das instituições. Exige mais detalhes sobre a coleta de dados, o volume e o formato. Insiste nas questões éticas e aconselha que se mencione uma pessoa/instituição responsável pela gestão dos dados.

Os dados e a publicação científica

A publicação dos dados de pesquisas está intrinsecamente associada aos artigos e, por isso, os editores científicos estão exigindo a disponibilização dos dados para aceitar o artigo. É mais um incentivo para que os pesquisadores disponibilizem os dados. A situação é nova para ambas as comunidades (a de autores e a de editores) e também para as relações entre os periódicos, os repositórios de dados etc.

Arend, Lange, Chen, Flemming, Hecht et al. mostraram o processamento dos dados, dos dados brutos até a publicação científica¹⁹, como se vê na Figura 7.

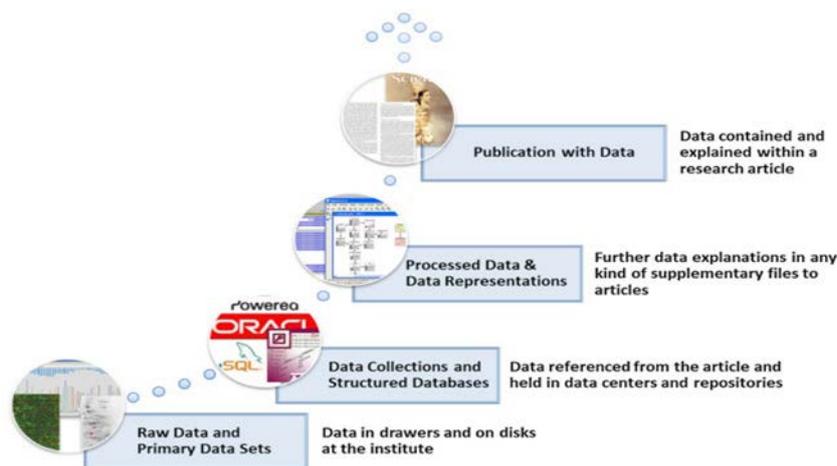


Figura 7 - Processamento dos dados: as etapas
Fonte: Arend et al. (2014).

Data paper ou artigo sobre dados de pesquisas

O *data paper* é um novo tipo de publicação que descreve um conjunto de dados científicos brutos e informações da descrição dos metadados, além do *Digital Object Identifier* – DOI do conjunto de dados disponibilizado em um repositório. O *data paper* é publicado em um periódico submetido à avaliação dos pares que informa a comunidade científica sobre a disponibilidade do conjunto de dados e o potencial para utilizações futuras. O *data paper* é muito interessante porque valoriza o conjunto de dados e também todo o grupo que trabalhou na sua elaboração. A submissão pode ser realizada por uma equipe que decidiu compartilhar dados.

Os *data papers* são publicados em periódicos especiais que se intitulam *data journals*, ou em periódicos clássicos que aceitam esse tipo de artigo, como por exemplo o *Ecology*, da Ecological Society of America.

Os *data journals* aceitam somente os *data papers* criados a partir de um modelo de acesso livre intitulado *auteur-pays*: os autores pagam uma taxa de processamento do artigo. Para isso, a editora Nature Group Publishing criou o periódico *Scientific Data*^{xxxix}. Outro é o *Geoscience Data Journal*^{xl}. Uma lista completa dos *data journals* está no trabalho de Candela, et al. realizado em 2015².

Política dos periódicos

Os periódicos e os editores estão envolvidos na difusão dos dados de pesquisas disponibilizados para reproduzir as experiências, aumentar as citações e a atratividade dos artigos, mas sobretudo para diminuir a fraude científica. Os periódicos permitem a difusão do material complementar já que os artigos têm limite de espaço e não tinham a indicação de onde os autores poderiam depositar os dados das pesquisas.

xxxix Disponível em: <<http://www.nature.com/sdata/>>.

xl Disponível em: <<http://www.geosciencedata.com/>>.

O projeto Journal Research Data Policy Bank (JoRD) do Joint Information Systems Committee (JISC) estudou 400 periódicos²⁰, os 200 mais citados e os 200 menos citados (segundo o Fator de Impacto), e mostrou que 50% dos periódicos não têm ainda nenhuma política em relação à abertura e ao depósito de dados. Quando têm alguma política, geralmente são heterogêneas em termos de exigências. Das 230 políticas analisadas segundo critérios de Piwowar e Chapman²¹, 76% foram definidas como fracas, 24% consideradas fortes, e entre elas apenas 8% exigem o depósito de dados para publicar o artigo, embora mais raras ainda sejam as que exigem a especificação dos repositórios. A integração do depósito de dados das pesquisas ao processo editorial varia, mas 51% dos casos exigem o comprovante de depósito dos dados em um repositório durante a submissão do artigo.

Desde 2013, esse processo vem evoluindo. Periódicos como as revistas *Nature* e a *Science* adotam políticas que exigem que os dados de pesquisas estejam abertos, estimulando outros periódicos a fazerem o mesmo. As instruções da National Information Standards Organization – NISO sobre material suplementar do artigo, sugerem colocar esse tipo de material em repositórios (Quadro 1) adequados que forneçam boa preservação e *link* bidirecional com o periódico²². Em setembro de 2014, a SciELO iniciou a disponibilização de dados como critério para admissão do artigo.

Citações de dados

A citação de dados de pesquisas estudada pelo Digital Curation Center²¹ mostrou que o reconhecimento do trabalho e de seus produtores estimula a publicação dos dados. O estabelecimento do *link* entre o artigo e os dados da pesquisa ajuda a aumentar a citação do artigo²², mas é ainda necessário um mecanismo mais robusto para a citação, já que permite também o aparecimento de novas medidas cientométricas.

A citação de dados deveria ser utilizada tanto pelas pessoas como pelo computador para identificar um objeto único e, se possível, outros subconjuntos de dados²³, como nas bases de dados bibliográficas.

A organização Datacite^{xli} apresenta um padrão mínimo para a citação. As recomendações são: criador, ano de publicação, título, editor e identificador.

O estudo de Piwowar e Vision²⁴ mostra que publicações com dados de pesquisas divulgados têm 9% a mais de citações, como por exemplo, na expressão de busca para *gene microarray*. Considerando os anos mais recentes, o número de citações aumentou 30%. A reutilização dos dados em outros artigos é cada vez mais importante.

O *Data Citation Index* – DCI^{xlii} é uma ferramenta da Thomson Reuters que indexa os conjuntos de dados de vários repositórios. Ele possui atualmente 2.626.528 conjuntos de dados²⁵, mas somente 12% desses conjuntos são citados. O DCI fornece as seguintes funções:

- Metadados homogeneizados;
- Funções de pesquisa avançadas;
- Número de citações;
- Integração com a plataforma Web of Knowledge.

xli Disponível em: <<http://schema.datacite.org/meta/kernel-3/index.html>>.

xlii Disponível em: <<http://thomsonreuters.com/data-citation-index/>>.

Conclusão

Na realidade, cientistas não têm tempo nem conhecimento para a gestão dos próprios dados. Um levantamento, realizado em 2010 com 1.300 cientistas mostrou que mais de 80% dos entrevistados gostariam de utilizar conjuntos de dados de outros cientistas. No entanto, em 2011, somente 36% afirmaram que disponibilizam os próprios dados de maneira amigável²⁶.

Apesar dos incentivos das instituições, universidades, agências de fomento e política dos periódicos, a maioria dos pesquisadores não disponibiliza os dados de pesquisas para os outros²⁷, embora reconheçam que compartilhar os dados contribui para o progresso científico.

O movimento de acesso aberto a dados de pesquisas tem algumas restrições que impedem o avanço da disponibilização, como a falta de informações sobre a valorização dos dados. Antes de propor a abertura de dados específicos, é preciso esclarecer quais dados serão abertos e quais serão embargados. Alguns dados são sensíveis como, por exemplo, informações confidenciais dos pacientes ou estratégias que não devem ser publicadas imediatamente. O embargo pode permitir aos pesquisadores explorar os próprios dados antes de disponibilizá-los.

Também é percebida uma certa falta de incentivo das instituições, muito embora as políticas sobre dados estabelecidas pelas universidades sejam consideradas cada vez mais importantes. Além disso, existe uma falta de profissionais especializados em curadoria e disponibilização de dados de pesquisas já que é uma atividade nova. Nesse caso, os bibliotecários têm um papel primordial e reconhecido pelas diferentes organizações na área.

Considerando o aumento do volume de dados de pesquisas, instituições pertinentes e órgãos governamentais deveriam também oferecer aos pesquisadores infraestrutura confiável, isto é, data centers para garantir a conservação dos dados de pesquisas ao longo do tempo nos repositórios.

O formato dos dados de pesquisas para o seu reuso é muito importante porque pode ampliar a utilização de formatos apropriados, como os formatos da Web Semântica e dos Linked Data definidos por Tim Berners-Lee²⁸. Os Linked Data permitem associar a interoperabilidade técnica dos dados²⁹ através da arquitetura da Web Semântica para estabelecer *links* entre uma variedade de fontes de dados.

Muitos cientistas estão envolvidos na *Open Science*, utilizando os dados produzidos por outros para comparar, completar ou processar os próprios dados. Hoje, esses dados estão sendo produzidos muito mais colaborativamente, e os estímulos em diferentes níveis mostram que o movimento de acesso aberto aos dados de pesquisas deve se ampliar.

Referências

1. Tukey JW. The future of data analysis. *Ann Math Statist* [Internet]. 1962 [cited 2016 Aug 02];33(1):1-67. Available from: <https://projecteuclid.org/euclid.aoms/1177704711>.
2. Candela L, Castelli D, Manghi P, Tani A. Data journals: a survey. *J Assoc Inf Sci* [Internet]. 2015 Jan [cited 2016 Aug 02];66(9):1747–62. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/asi.23358/abstract>
3. Sayão LF, Sales LF. Dados abertos de pesquisa: ampliando o conceito de acesso livre. *RECIIS: Rev Eletron de Comun Inf Inov Saúde* [Internet]. 2014 June [citado em 2016 Aug 02];8(2):76-92. Disponível em: <http://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>.
4. University of Melbourne. Management of research data and records policy (MPF1242) [Internet]. Parkville; 2013 Nov [cited 2016 Aug 02]. Available from: <https://policy.unimelb.edu.au/MPF1242#section-5>.
5. Australian National Data Service. What is research data? [Internet]. Caulfield East; [2015?]. Available from: <http://ands.org.au/guides/what-is-research-data.html>.

6. Hey T, Tansley S, Tolle K. The fourth paradigm: data-intensive scientific discovery. Redmond: Microsoft Research; 2009.
7. Hey T, Tansley S, Tolle K. Tansley Jim Gray on eScience: a transformed scientific method. Based on the transcript of a talk given by Jim Gray to the NRC-CSTB in Mountain View, CA, on January 11, 2007 [Internet]. In: _____. The fourth paradigm: data-intensive scientific discovery. Redmond: Microsoft Research; 2009 [cited 2016 Aug 02]. p. XVII-XXXI. Available from: <http://language-log.ldc.upenn.edu/myl/JimGrayOnE-Science.pdf>.
8. Nicol A, Caruso J, Archambault E. Open data access policies and strategies in the european research area and beyond. Québec: Science-Matrix, 2013 [cited 2016 Aug 02]. Available from: http://www.science-matrix.com/pdf/SM_EC_OA_Data.pdf.
9. Organisation for Economic Co-Operation and Development. OECD principles and guidelines for access to research data from public funding [Internet]. Paris; 2007 [cited 2016 Aug 02]. Available from: <https://www.oecd.org/sti/sci-tech/38500813.pdf>.
10. Digital Curation Center. Overview of funders' data policies [Internet]. Edinburgh; 2014 [cited 2016 Aug 02]. Available from: <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>.
11. Valen D, Blanchat K. Overview of OSTP responses [Internet]. London; 2015 [cited 2016 Aug 02]. Available from: <http://dx.doi.org/10.6084/m9.figshare.1367165>.
12. Dietrich D, Adamus T, Miner A, Steinhart G. De-Mystifying the Data Management Requirements of Research Funders [Internet]. Science and Technology Librarianship, 2012 [cited 2016 Aug 02]. Available from: <http://www.istl.org/12-summer/refereed1.html>.
13. SciELO. Critérios, política e procedimentos para a admissão e a permanência de periódicos científicos na coleção SciELO Brasil [Internet]. São Paulo; 2014 [citado em 2016 Aug 02]. Disponível em: http://www.scielo.br/avaliacao/20141003NovosCriterios_SciELO_Brasil.pdf.
14. IBICT. Promovendo o uso do repositório de dados de pesquisa DATAVERSE. O Ibict e a UnB dão início à primeira rede de repositórios de pesquisa DATAVERSE no Brasil [Internet]. Brasília; 2015 [citado em 2016 Aug 02]. Disponível em: <http://cariniana.ibict.br/index.php/noticias/193-a-cariniana-promove-o-uso-de-repositorio-de-dados-de-pesquisa-dataverse>.
15. Sayão LF, Sales LF. Guia de gestão de dados de pesquisa: Centro de Informações Nucleares [Internet]. Rio de Janeiro; 2015 [citado em 2016 Aug 02]. Disponível em: <http://www.cnen.gov.br/component/content/article?id=160>.
16. OpenAire. What is the research data pilot? [Internet]. 2016 [citado em 2016 Aug 02]. Available from: <https://www.openaire.eu/opendatapilot>.
17. G8 open data charter [Internet]. Fermanagh; 2013 [cited 2016 Aug 02]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf.
18. Digital Curation Center. Checklist for a Data Management Plan [Internet]. v.4.0. Edinburgh; 2013 [cited 2016 Aug 02]. Available from: <http://www.dcc.ac.uk/resources/data-management-plans>.
19. Arend D, Lange M, Chen J, Colmsee C, Flemming S, Hecht D, et al. e!DAL - a framework to store, share and publish research data [Internet]. BMC Bioinformatic. 2014 [cited 2016 Aug 02];15(214). Available from: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-214>.
20. Sturges P, Bamkin M, Anders JHS, Hubbard B, Hussain A, Heeley M. Research data sharing: developing a stakeholder-driven model for journal policies [Internet]. J Assoc Inf Sci. 2015 [cited 2016 Aug 02];2445-55. Available from: <http://onlinelibrary.wiley.com/doi/10.1002/asi.23336/epdf>.
21. Ball A, Duke M. Data citation and linking'. DCC briefing papers [Internet]. Edinburgh: Digital Curation Centre; [cited 2016 Aug 02] 2012. Available from: <http://www.dcc.ac.uk/resources/briefing-papers/>.
22. Piwowar H, Chapman W. Public sharing of research datasets: a pilot study of associations [Internet]. J Informetr. 2010 Apr [cited 2016 Aug 02];4(2):148-56. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039489/>.
23. National Information Standards Organization. National Federation of Advanced Information Services. NISO RP-15-3013 recommended practices for online supplemental journal article materials [Internet]. Baltimore; 2013 Jan [cited 2016 Aug 02]. Available from: http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf.

24. Piwowar H, Vision T. Data reuse and the open data citation advantage [Internet]. [Peer J](#). 2013 Oct [cited 2016 Aug 02];1:e175. Available from: <https://peerj.com/articles/175/>.
25. Torres-Salinas D, Jimenez-Contreras E, Robinson-Garcia N. How many citations are there in the Data Citation Index? [Internet]. In: STI Conference Leiden; 2014 Sept 3-5; ENID, CWTS; 2014 [cited 2016 Aug 02]. Available from: <https://arxiv.org/ftp/arxiv/papers/1409/1409.0753.pdf>.
26. Tenopir C, Allard S, Douglass K, Aydinoglu A, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions [Internet]. PLoS ONE
27. 2011 [cited 2016 Aug 02];6(6):1-21. Available from: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101>.
28. Fecher B, Friesike S, Hebing M. What drives academic data sharing? [Internet]. PLoS ONE 2015 [cited 2016 Aug 02];10(2):e0118053. Available from: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053>.
29. Berners-Lee T. The year open data went worldwide [video]. New York: Ted Ideas Worth Spreading, 2010 Jan [cited 2016 Aug 02]. 5:26min color. son. Available from: http://www.ted.com/talks/tim_berniers_lee_the_year_open_data_went_worldwide.
30. 7 Things you should know about...linked data [Internet]. In: Coar-Sparc Conference 2015 [cited 2016 Aug 02]. Available from: <https://www.coar-repositories.org/activities/repository-observatory/second-edition-linked-open-data/7-things-you-should-know-about-open-data/>.
31. DOI. In: Wikipedia: a enciclopedia livre. [S.l.: s.n.], [2015?] [citado em 2016 July 28]. Available from: https://pt.wikipedia.org/wiki/Digital_object_identifier.