

# Explorando o Google Data Search para zika, dengue e chikungunya

## **Maria Simone de Menezes Alencar**

Doutora em Engenharia de Processos Químicos e Bioquímicos pela Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ – Brasil. Professora da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) - Rio de Janeiro, RJ – Brasil.

<http://lattes.cnpq.br/1933365741854449>

E-mail: [simone.alencar@unirio.br](mailto:simone.alencar@unirio.br)

## **Jorge Lima de Magalhães**

Pós-Doutorado pela Université de Provence Aix Marseille I (Aix-Marceille I) - França. Doutor em Engenharia de Processos Químicos e Bioquímicos pela Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ – Brasil. Pesquisador em Saúde Pública e Coordenador do PPG Profissional em Gestão, Pesquisa e Desenvolvimento na Indústria Farmacêutica da Fundação Oswaldo Cruz (FIOCRUZ) - Rio de Janeiro, RJ - Brasil

<http://lattes.cnpq.br/9829199474735249>

E-mail: [jorge.magalhaes@far.fiocruz.br](mailto:jorge.magalhaes@far.fiocruz.br)

## **Suzanne de Oliveira Rodrigues Schumacher**

Mestrado profissional em Propriedade Intelectual e Inovação pelo Instituto Nacional da Propriedade Industrial (INPI) - Rio de Janeiro, RJ - Brasil. Bolsista e pesquisadora do Sistema de Informação sobre a Indústria Química (SIQUIM/EQ/UFRJ) - Rio de Janeiro, RJ - Brasil.

<http://lattes.cnpq.br/6933358057465305>

E-mail: [suzanne@eq.ufjf.br](mailto:suzanne@eq.ufjf.br)

## **Juliana de Simone Morais**

Graduação em Química Industrial pela Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ - Brasil

<http://lattes.cnpq.br/3939010423789523>

E-mail: [ju.s.morais@eq.ufjf.br](mailto:ju.s.morais@eq.ufjf.br)

## **Mateus Pinheiro Ramos**

Graduando em Engenharia Química pela Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ – Brasil. Pesquisador Estagiário do Sistema de Informação sobre a Indústria Química (SIQUIM/EQ/UFRJ) - Rio de Janeiro, RJ - Brasil.

<http://lattes.cnpq.br/4198948795429519>

E-mail: [mateusramos@eq.ufjf.br](mailto:mateusramos@eq.ufjf.br)

## **Adelaide Maria de Souza Antunes**

Pós-Doutorado pela Instituto Francês de Petróleo (IFP) - França. Doutora em Engenharia Química pelo Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE/UFRJ) - Rio de Janeiro, RJ - Brasil. Professora do Mestrado Profissional e do Doutorado Profissional em Propriedade Intelectual e Inovação do Instituto Nacional da Propriedade Industrial (INPI) - Rio de Janeiro, RJ – Brasil. Professora da Universidade Federal do Rio de Janeiro (UFRJ) - Rio de Janeiro, RJ - Brasil.

<http://lattes.cnpq.br/5168823546200073>

E-mail: [adelaide@eq.ufjf.br](mailto:adelaide@eq.ufjf.br)

Data de submissão: 26/09/2019. Data de aprovação no ConfOA: 12/06/2019. Data de publicação:

## **RESUMO**

O trabalho visa explorar o Google Data Search, ferramenta instrumento de apoio à Ciência Aberta, composto por vários provedores de diferentes assuntos. O estudo teve por base a área de saúde pública com busca de dados sobre as doenças negligenciadas zika, dengue e chikungunya. O levantamento obteve 75 conjuntos de dados que foram tratados em planilha Excel por título; nome do provedor com respectivo link; nº de artigos que citam os dados; DOI; datas da criação, atualização, publicação; instituição fornecedora dos dados; autores, licença e formato de download. Após análise, foram criadas mais duas variáveis indexando a região geográfica e tipologia dos dados. Os resultados mostraram que experimentos, casos e modelagem estatística são as tipologias mais frequentes, 45% dos conjuntos de dados são citados em artigos científicos e a região que tem o maior número de conjunto de dados sobre os temas é o Brasil. Os metadados são originados de 13 diferentes provedores. Conclui-se sobre o potencial que o Google Data Search tem para estimular o reuso dos dados abertos de pesquisa, contribuindo para o avanço da ciência e maximização dos investimentos aplicados em pesquisas.

**Palavras-chave:** Dados de pesquisa. Reuso. Ciência Aberta. Dados abertos de pesquisa.

## **A Google Data Search overview for zika, dengue and chikungunya**

### **ABSTRACT**

*The paper aims to explore Google Data Search – an open science support tool composed of several providers of different subjects. The study was based on public health with data on neglected diseases such as zika, dengue and chikungunya. The work obtained 75 datasets that were treated in Excel spreadsheet by: title; provider name with its link; at the. of articles citing the data; DOI; dates of creation, update, publication; institution providing the data; authors, license and download format. After analysis, two more variables were created indexing the geographic region and data typology. The results showed that experiments, cases and statistical modeling are the most frequent typologies, 45% of the datasets are cited in scientific articles and the region with the largest number of datasets on the themes is Brazil. Metadata originates from 13 different providers. It concludes about the potential that Google Data Search must stimulate the reuse of open search data, contributing to the advancement of science and maximizing investments in search.*

**Keywords:** Research Data. Reuse. Open Science. Open Research Data.

## **Explorando la búsqueda de datos de Google para zika, dengue y chikungunya**

### **RESUMEN**

*El trabajo tiene como objetivo explorar Google Data Search, una herramienta abierta de apoyo científico, compuesta por varios proveedores de diferentes temas. El estudio se basó en la salud pública con datos sobre enfermedades olvidadas zika, dengue y chikungunya. La encuesta obtuvo 75 conjuntos de datos que fueron tratados en una hoja de cálculo Excel por: título; nombre del proveedor con su enlace; no de artículos que citan los datos; DOI; fechas de creación, actualización, publicación; institución que proporciona los datos; autores, licencia y formato de descarga. Después del análisis, se crearon dos variables más que indexan la región geográfica y la tipología de datos. Los resultados mostraron que los experimentos, los casos y el modelado estadístico son las tipologías más frecuentes, el 45% de los conjuntos de datos se citan en artículos científicos y la región con el mayor número de conjuntos de datos sobre los temas es Brasil. Los metadatos se originan en 13 proveedores diferentes. Concluye sobre el potencial que tiene la búsqueda de datos de Google para estimular la reutilización de datos de búsqueda abiertos, contribuyendo al avance de la ciencia y maximizando las inversiones en búsqueda.*

**Palabras clave:** Datos de investigación. Reutilización. Ciencia abierta. Datos abiertos de investigación.

## INTRODUÇÃO

As transformações recentes por que passa a construção da forma de se fazer a ciência estão exigindo competências relacionadas ao compartilhamento e ao reúso de dados de pesquisa, exigindo novos processos, métodos e ferramentas de apoio aos pesquisadores, haja vista o crescente número de bytes novos diários adicionados à Web: 2,5 quintilhões ( $2,5 \times 10^{18}$ ) de bytes (HUBERMAN, 2012; MAGALHAES *et al.*, 2018).

O Google Data Search visa apoiar o desenvolvimento de um ecossistema de dados abertos na Internet. Segundo definição da Open Knowledge International, dados são considerados abertos quando qualquer cidadão pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito, no máximo, a exigências de atribuição da fonte e compartilhamento pelas mesmas regras (OPEN DATA HANDBOOK, 2019).

Embora esse conceito seja amplo o suficiente para abarcar qualquer tipo de dado, cada vez mais vem se destacando na comunidade científica os dados abertos de pesquisa. Segundo a Organisation for Economic Co-operation and Development (2007), os dados de pesquisa podem ser definidos como registros factuais usados como fontes primárias de pesquisa científica e que são aceitos na comunidade científica como necessários para validar resultados de pesquisa. Na mesma linha, recentemente Semeler e Pinto (2019) realizaram uma interessante revisão sobre os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados, definindo-os como o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica.

Como um dos pilares da Ciência Aberta, os dados de pesquisa são frequentemente considerados o resultado mais valioso de muitos projetos de investigação, pois permitem o desenvolvimento de novas descobertas teóricas ou aplicadas. São produtos heterogêneos e contextualizados conforme as disciplinas do conhecimento.

Segundo Pampel e Tiessen (2014), os dados de pesquisa podem ser disponíveis de forma aberta mediante três estratégias: (a) publicação como um objeto de informação independente, através de um repositório; (b) publicação dentro de um documento textual, o chamado *data-paper* ou jornal de dados; (c) publicação como anexo de um artigo, também chamado de publicação ampliada.

Seguindo a primeira estratégia, os dados de pesquisa podem estar agregados nos repositórios institucionais, disciplinares, multidisciplinares e de projetos específicos. Essa tipologia foi apresentada por Pampel e colaboradores (2013) ao discorrer sobre o re3data.org—Registry of Research Data Repositories, um “repositório de repositórios” lançado em 2012 que atualmente agrega mais de dois mil repositórios das mais diversas áreas e diferentes países. Importante destacar que não há uma busca integrada, ante a heterogeneidade dos dados que compõem cada um dos conjuntos de dados e a inexistência de um padrão de metadados e interoperabilidade comum.

Observa-se então a dificuldade que um pesquisador encontra de localizar dados de pesquisa, já que eles podem estar em diversos locais. O reúso dos dados é dificultado, pois identificar onde está o conjunto de dados que contém as informações desejadas é um processo longo: é necessário identificar se há um repositório que agregue informações sobre o tema desejado, para então entrar no repositório e executar a busca de acordo com os campos que ele ofereça. Além disso, identificar e buscar nos jornais de dados e nas publicações ampliadas de sua área de conhecimento.

Seguindo a linha de produtos da família Google – páginas, imagens, notícias, vídeos etc. – de busca de informação com uma interface simples, e visando facilitar a localização de conjunto de dados, foi lançado em setembro de 2018 o Google Data Search. Utiliza tecnologia para integração de dados empresariais com o sistema Goods, e seu objetivo é disponibilizar informações de diferentes provedores de dados agregando valor com a adição de metadados estruturados, usando um padrão aberto (HALEVY *et al.*, 2016). Os metadados especificam as propriedades explícitas de cada conjunto de dados: seu nome e descrição, dados de criação, atualização, informações de proveniência e outros recursos disponíveis no Google, tais como publicações ampliadas por meio do Google Acadêmico (CHAPMAN *et al.*, 2019). Aparentemente irá facilitar o acesso a conjunto de dados abertos disponíveis na internet, sendo permitido aos fornecedores de dados publicar seus dados, com orientações sobre padrões de marcação e metadados a serem utilizados.

Visando explorar essa ferramenta, foi escolhido um tema relevante para a saúde pública, uma vez que na última década o país vem deparando com três doenças tropicais negligenciadas (DN), cuja característica é ser transmitida através da picada de mosquitos do gênero *Aedes*. Eles funcionam como transmissores ou vetores das doenças chikungunya, dengue e zika. As três, consideradas doenças negligenciadas (DN), têm sido chamadas de ameaça tríplice, em razão de possuírem o mesmo gênero do causador da doença e apresentarem epidemiologia parecida.

## **METODOLOGIA**

A busca foi realizada no site Google Dataset Search utilizando-se a estratégia “dengue AND zika AND chikungunya”.

A partir do levantamento obtido, foi criada uma planilha Excel com os metadados obtidos na ferramenta de busca. Foram armazenados os seguintes metadados: título, provedor de dados, link do repositório, número de artigos que citam o conjunto de dados, link para os artigos que citam o conjunto de dados, DOI, data de criação, data de atualização, data de publicação, instituição fornecedora dos dados, autores, licença, formato de download disponibilizado pelo provedor e descrição. Quando algum metadado estava não disponível, foi incluída a informação “nd” na planilha.

Foram gerados mais dois campos adicionais: região geográfica e tipologia dos dados, extraídos a partir da observação das informações contidas no título ou da descrição do conjunto de dados. O campo foi chamado região, pois pode ser concernente a mais de um país ou área geográfica a que se referem os dados.

A partir da extração automática, através do software VantagePoint<sup>®</sup>, dos termos contidos no título e na descrição dos conjuntos de dados, foi feita uma nuvem de termos, utilizando a ferramenta WordClouds, para se obter uma aproximação dos temas envolvidos nos resultados da busca.

## **ANÁLISE E DISCUSSÃO DOS RESULTADOS**

A busca foi realizada em 12 de março de 2019 e foram encontrados 75 conjuntos de dados com a estratégia definida. A título ilustrativo, a figura 1 apresenta uma tela do resultado de busca.



A primeira análise realizada refere-se à construção da nuvem que foi feita a partir das palavras-chave contidas no título e na descrição apresentada na figura 2.

Cabe ressaltar que ao considerar a nuvem de palavras no campo descrição, conclui-se que somente o gênero comum do mosquito causador do trio de doenças (aedes) aparece em 19% dos resultados encontrados, ou 14 repositórios, como elencam trabalhos contendo a tríplice ameaça. Já ao filtrar pelo gênero “aedes aegypti”, o resultado alcança 47% do total, ou seja, 35 repositórios. Não obstante, o gênero “aedes albopictus” e suas outras composições de palavras (dengue, zika ou chikungunya) aparecem nos demais 53% dos resultados.

A segunda análise refere-se à completude das informações disponibilizadas, apresentada na tabela 1.

Tabela 1 – Campos e completude de disponibilidade

Campo	% Disponível
Título	100%
Nome do provedor de dados	100%
Link do provedor	100%
Nº de artigos que citam o dado	45%
DOI	81%
Conjunto de dados criado	76%
Conjunto de dados atualizado	56%
Conjunto de dados publicado	63%
Instituição fornecedora dos dados	60%
Autores	85%
Licença	84%
Formatos de download disponibilizados pelos provedores	12%
Descrição	100%
Região geográfica	59%
Tipologia dos dados	96%

Fonte: Elaboração dos autores.

Os dados coletados estão presentes em 14 diferentes provedores, sendo de maior frequência o Figshare (31 conjuntos de dados), conforme pode ser visto na tabela 2. Alguns conjuntos de dados estão registrados em mais de um provedor, e por isso a soma do número de provedores é maior do que o número total de registros localizados na busca. A seguir apresentam-se informações sobre os provedores, obtidas nos respectivos sites.

Tabela 2 – Provedores de dados e respectivos números de conjuntos de dados

Nome e endereço do provedor de dados	No de conj. de dados
Figshare ( <a href="https://figshare.com/">https://figshare.com/</a> )	31
OpenAIRE Explore ( <a href="https://explore.openaire.eu/">https://explore.openaire.eu/</a> )	25
Dryad Digital Repository ( <a href="https://datadryad.org">https://datadryad.org</a> )	6
DataCite ( <a href="https://search.datacite.org/">https://search.datacite.org/</a> )	4
Portal de Dados Abertos - Governo colombiano ( <a href="https://www.datos.gov.co">https://www.datos.gov.co</a> )	4
Data.world ( <a href="https://data.world/">https://data.world/</a> )	2
Kaggle ( <a href="https://www.kaggle.com/">https://www.kaggle.com/</a> )	2
City of Denton Open Data ( <a href="http://data.cityofdenton.com/">http://data.cityofdenton.com/</a> )	1
Datamed ( <a href="https://datamed.org/">https://datamed.org/</a> )	1
Humanitarian Data Exchange ( <a href="https://data.humdata.org/">https://data.humdata.org/</a> )	1
NARCIS ( <a href="https://www.narcis.nl/">https://www.narcis.nl/</a> )	1
Research Data Australia ( <a href="https://researchdata.andis.org.au/">https://researchdata.andis.org.au/</a> )	1
Zenodo ( <a href="https://zenodo.org/">https://zenodo.org/</a> )	1

Fonte: Elaboração dos autores.

Dentre os provedores de dados, observam-se repositórios científicos abertos, tais como o Figshare, Dryad e Zenodo.

O Figshare foi lançado em 2011 e pertence à Digital Science, empresa de tecnologia localizada em Londres. É um repositório científico aberto que hospeda dados e resultados gerados por pesquisas das mais diversas áreas.

O Dryad Digital Repository é um repositório digital que se declara “com curadoria que torna os dados de pesquisa detectáveis, livremente reutilizáveis e citáveis” (tradução nossa).

O Zenodo, embora com apenas um resultado nessa pesquisa, configura-se como um repositório científico de conjuntos de dados, softwares, relatórios e demais artefatos digitais gerados a partir de resultados de pesquisa de diversas áreas do conhecimento. Foi desenvolvido no âmbito do programa europeu OpenAIRE e operado pelo CERN1. O OpenAIRE (Open Access Infrastructure for Research in Europe) foi um projeto lançado em 2009 pelo programa FP7 (7th Framework Programme for Research and Technological Development) da Comissão Europeia, cujo objetivo foi implementar o acesso aberto na Europa.

Também no escopo desse programa está o provedor com o segundo maior número de resultados nessa busca, o OpenAIRE Explorer, interface desenvolvida recentemente pelo projeto OpenAIRE visando coletar dados de pesquisa de diferentes fontes para tratá-los e extrair conhecimento em uma plataforma de Business Intelligence. Essa plataforma de BI é o OpenAIRE Monitor, projetada para ajudar financiadores e iniciativas de pesquisa a rastrear, analisar, monitorar e documentar seus dados de pesquisa (VOS et al., 2019).

O Datacite é um consórcio internacional fundado em 2009, com o objetivo de melhorar a citação de dados mediante o uso de identificadores persistentes associados aos dados e conjuntos de dados de pesquisa. Embora não seja exatamente um provedor, estabelece acesso a dados de pesquisa através de sua interface de busca.

Dois provedores se caracterizam como fontes de dados governamentais: o Portal de Dados abertos do governo colombiano, e o City of Denton Open Data, sendo o último de uma cidade do Texas, nos Estados Unidos. O data.world é uma empresa de análise de dados. O DataMed é um protótipo de mecanismo de pesquisa de dados biomédicos, cujo objetivo é descobrir conjuntos de dados entre repositórios ou agregadores de dados.

O Kaggle é uma plataforma para competições de Data Science. A plataforma foi fundada em 2010, e em 2017 foi adquirida pelo Google.

O Humanitarian Data Exchange (HDX) é uma plataforma aberta para compartilhar dados entre crises e organizações. Lançado em julho de 2014, o objetivo do HDX é facilitar a localização e o uso de dados humanitários.

Dois provedores são repositórios nacionais: o Narcis refere-se à produção científica da Holanda, e o Research Data Australia é um portal da Web para descobrir conjuntos de dados produzidos por ou relevantes para pesquisadores australianos. Observa-se, portanto, uma gama de fontes de informação coletadas pelo Google Data Search, desde fontes acadêmicas, até comerciais e governamentais.

Ao avaliar a recuperação dos dados pela ótica da estratégia de publicações ampliadas, observa-se que dos 75 conjuntos de dados, apenas 45% são citados em artigos científicos, variando entre 2 e 82 o número de citações. Os resultados apresentados pela ferramenta não permitiram analisar a publicação em jornais de dados.

Os campos referentes às datas de criação, publicação e atualização ainda não permitem uma análise mais apurada, pois os dados de pesquisa começaram a ser disponibilizados em 2016, não havendo intervalo temporal suficiente para análises consistentes. As instituições fornecedoras dos dados referem-se a quem está divulgando a informação.

Embora em muitos casos (n=30) essa informação não fosse divulgada, dentre os dados disponíveis observou-se de maior a presença de revista científica *PLOS Neglected Tropical Diseases* (n=8) e uma instituição que tem como objetivo a promoção do conhecimento científico, a Royal Society (n=8). Observou-se também a presença de provedores, como Dryad Digital Repository, Figshare e DataCite. Outras instituições elencadas foram tão diversas como: Anonymous citizen scientists collaborating through the Mosquito Alert platform, Atlas of Living Australia, Centre d'Estudis Avançats de Blanes (CEAB-CSIC), City of Denton GIS Technology Services, Frontiers Media, Karger Publishers, PLOS Medicine, Pompeu Fabra University, Universidad de Murcia, University of the Balearic Islands e a University of Zaragoza.

Pode-se supor que esse metadado refere-se a diferentes informações, desde o nome do periódico em que o artigo foi publicado, como ao repositório, bem como ao provedor primário de dados. Tal inferência indica que é preciso um padrão de metadados mais consistente para detalhar a que se refere “instituição fornecedora dos dados”.

Em relação à autoria, foram identificados 442 diferentes autores compondo 53 conjuntos de coautores. Como podem ocorrer erros de grafia, esses dados são aproximados. Observou-se a existência de conjuntos de dados com um autor e até 22 autores.

As licenças apresentadas são licenças abertas, mas em diversas formas (Creative Commons, Open Access, CC-0, public domain), indicando falta de padronização dos campos. A de maior frequência é a Creative Commons, organização que oferece licenças de direitos autorais gratuitas (CREATIVE COMMONS, 2019). Esse tipo de licença permite que terceiros façam uso de obras sem infringir as leis de propriedade intelectual, facilitando o compartilhamento do conhecimento (UnB, 2019).

Os formatos de download oferecidos são bastante diversos: csv, json, rdxml, rssxml, xml, zip, html, xlsx, dentre outros. No entanto, observam-se formatos proprietários, tal como xlsx, mas os dados devem ser divulgados em formato aberto, “para que não exista dependência tecnológica” (MANUAL DOS DADOS ABERTOS, 2011, p.24).

O campo descrição traz mais informações sobre os dados. Em alguns conjuntos é apenas uma linha e em outros é bem minucioso, fornecendo informações detalhadas sobre a pesquisa, figuras e referências.

Em relação à região geográfica, essa informação foi obtida a partir da observação do título e descrição dos registros recuperados, e se refere ao conjunto de dados, não necessariamente ao produtor dos dados. Em 31 conjuntos de dados essa informação não estava disponível, e o país com maior frequência foi o Brasil (n=5), seguido por dados globais (n=5) e Colômbia e Estados Unidos com n=3. Dada a importância da tríplice ameaça, tema da pesquisa no Brasil, sua presença pode ser explicada.

Quanto à tipologia dos dados, foram observados: experimentos (n=38), casos (n=21), modelagem (n=12), características clínicas (n=7), incidência do mosquito (n=3) e dados de cenários (n=2). Nota-se uma variedade de pesquisas sendo realizadas, utilizando metodologias observacionais, experimentais e computacionais.

Em recente publicação, Sales e Sayão (2019) sugeriram uma taxonomia para dados de pesquisa, propondo que o estudo de suas propriedades em cada domínio do conhecimento é uma condição para a construção de critérios de seleção para facilitar o reuso dos dados.

## CONCLUSÕES

Com apenas um ano de existência, a ferramenta apresentou resultados amplos e consistentes, de fontes variadas, desde repositórios científicos genéricos a temáticos, provedores institucionais, governamentais e comerciais. O Google Data Search mostra-se assim uma ferramenta com potencial para estimular o ecossistema da Ciência Aberta em seu pilar dos dados de pesquisa abertos.

No entanto, para que eles possam ser efetivamente recuperados, é necessário que sigam padrões das comunidades de pesquisa. Nesse sentido, os princípios FAIR, acrônimo de *Findable, Accessible, Interoperable e Reusable*, precisam ser adotados na fonte, ou seja, na indexação dos dados em sua origem.

Com o aumento da recuperação de dados de pesquisa por meio dessa ferramenta, espera-se que aumente o seu reuso. A literatura é vasta em indicar as vantagens no reuso de dados de pesquisa, desde o aumento das citações até a reprodutibilidade da pesquisa. Porém em termos mais amplos, pode-se afirmar que o desenvolvimento de ferramentas como essa, que facilitam o reuso de dados, contribui de maneira significativa para que a ciência avance e maximize os investimentos aplicados em programas de pesquisa, desenvolvimento e inovação.

---

## REFERÊNCIAS

CHAPMAN, A. *et al.* Dataset search: a survey. *arXiv preprint arXiv:1901.00735*, 2019.

CREATIVE COMMONS. O que é Creative Commons? [S.l.], 2019. Disponível em: <https://br.creativecommons.org/sobre/>. Acesso em: 20 de março de 2019.

HALEVY, A. *et al.* Goods: Organizing google's datasets. In: INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 2016, San Francisco. *Proceedings [...]*. New York: ACM, 2016. p. 795-806.

OPEN DATA HANDBOOK. *What is open data*. [S.l.], 2019. Disponível em: <http://opendatahandbook.org/>. Acesso em: 25 de ago. 2019.

HUBERMAN, B. A. Sociology of Science: Big data deserve a bigger audience. *Nature*, [S.l.], v.482, n.7385, p. 303-308, 16 fev.2012.

MAGALHÃES, J. L. *et al.* The Contribution of Information Science Through Scientific and Technological Knowledge in Intellectual Property. In: JAMIL, G. L. (org.). *Advances in Business Information Systems and Analytics*. Pennsylvania: IGI Global, 2018. p. 265-279.

MANUAL dos dados abertos: desenvolvedores. São Paulo: Comitê Gestor da Internet no Brasil, 2011. Disponível em: [http://www.w3c.br/pub/Materiais/PublicacoesW3C/manual\\_dados\\_abertos\\_desenvolvedores\\_web.pdf](http://www.w3c.br/pub/Materiais/PublicacoesW3C/manual_dados_abertos_desenvolvedores_web.pdf). Acesso em: 20 de março de 2019.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *OECD principles and guidelines for access to research data from public funding*. Paris: OECD Publications, 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 26 ago. 2019.

PAMPEL, H. *et al.* Making research data repositories visible: the re3data.org registry. *PLoS one*, [S.l.], v. 8, n. 11, p. e78080, 2013.

PAMPEL, H.; DALLMEIER-TIESSEN, S. Open research data: From vision to practice. In: BARTLING, S.; FRIESIKE, S (ed.). *Opening science*. Cham: Springer, 2014. p. 213-224.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. *Conhecimento em ação*, Rio de Janeiro, v. 4, n. 1, jan./jun., 2019. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337/14573>. Acesso em: 30 ago. 2019.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. *Ciência da Informação*, Brasília, v. 48, n. 1, 2019.

UNIVERSIDADE DE BRASÍLIA (UNB). *Repositórios de Acesso Aberto e as Licenças Creative Commons*. Brasília: [s.n.], 2019. Disponível em: [http://repositorio.unb.br/documentos/Creative\\_Commons.pdf](http://repositorio.unb.br/documentos/Creative_Commons.pdf). Acesso em: 20 de março de 2019.

VOS, A. *et al.* OPENAIRE Monitor Dashboard: an added value service for funders and research initiatives. [S.l.], 2019. Disponível em: <https://www.opensciencefair.eu/posters-2019/openaire-monitor-dashboard-an-added-value-service-for-funders-and-research-initiatives>. Acesso em: 25 set. 2019.