

A construção do repositório de dados da UFPB: a experiência com o dataset de Arboviroses

The construction of the ufpb data repository: the experience with the Arboviruses Dataset

Pollianna Marys de Souza e Silva

Universidade Federal da Paraíba
pollianna_marys@hotmail.com

Sandra de Albuquerque Siebra

Universidade Federal de Pernambuco

Resumo

Repositório de Dados (RD) são sistemas digitais de informação que armazenam dados provenientes de pesquisas científicas, a fim de colaborar com o seu acesso, uso, reuso e preservação. Nesse contexto, este artigo objetiva relatar a experiência de construção do RD da UFPB. Esta foi uma pesquisa-ação, qualitativa e descritiva. Como resultados iniciais se obteve a instalação da plataforma Dataverse e criação do RD da UFPB, no ano de 2018, tendo como conjunto de dados inicial um dataset construído a partir da extração de mais de um milhão de posts da rede social Twitter, de outubro de 2017 a março de 2018, sobre as arboviroses, que é um conjunto de patologias formado principalmente pela Zika, Dengue e Chikungunya. Com esse dataset espera-se disponibilizar dados brutos que possam servir para pesquisadores de várias áreas estudarem a disseminação das doenças, do que se fala sobre elas, mapear seus agravos e surtos, entre outros.

Palavras-chave: Repositório de Dados; Dataverse; Dataset Arboviroses; Curadoria de Dados; Ciência Aberta.

Abstract

Data Repository (RD) are digital information systems that store data from scientific research in order to collaborate with its access, use, reuse and preservation. In this context, this article aims to report on the experience of building the UFPB's data repository. This was an action research, qualitative and descriptive. As initial results we obtained the installation of the Dataverse platform and creation of the RD of the UFPB, in the year of 2018, having as dataset a dataset constructed from the extraction of more than one million posts of the social network Twitter, of October of 2017 to March 2018, on arboviruses, which is a set of pathologies formed mainly by Zika, Dengue and Chikungunya. With this dataset, it is expected to provide raw data that can be used for researchers from various areas to study the spread of diseases, what is being said about them, map diseases and their outbreaks, among others.

Keywords: Data Repository; Dataverse; Arboviruses Dataset; Data Curation; Open Science.

I INTRODUÇÃO

A crise dos periódicos, o paradigma tecnológico, a criação do arXiv e a realização das conferências de Budapeste, Berlin; Bethesda (3B) foram fatores fundamentais para a emergência do acesso livre e para as consequentes alterações provocadas nos modos de produção e comunicação da ciência.

Nessa perspectiva, o primeiro repositório de acesso aberto foi concebido pelo físico americano Paul Ginsparg, do Los Alamos National Laboratory, situado na cidade de Los Alamos, New Mexico, Estados Unidos. Denominado arXiv, surgiu como uma alternativa ao sistema formal de comunicação científica e se constituiu um repositório de *preprints* eletrônicos na área da Física, passando a incluir, posteriormente, disciplinas como a Astronomia, a Matemática, as Ciências da Computação, a Ciência Não-Linear, a Biologia Quantitativa e, mais recentemente, a Estatística (SHINTAKU; MEIRELES, 2010; AUTRAN; BORGES, 2014).

O movimento em favor do acesso aberto, conhecido como Open Access, surgiu a partir da crise dos periódicos e permitiu a democratização do acesso à informação, sendo os repositórios digitais uma das primeiras plataformas digitais de acesso aberto (SILVA JÚNIOR; BORGES, 2014). Estes são ambientes digitais que possibilitam reunir dados e informações de cunho científico, administrativo, técnico, artístico, cultural, entre outros, cuja função principal é promover a visibilidade de seus objetos digitais, preservando-os por meio do gerenciamento de informação (ABREU; VIDOTTI, 2016). Eles podem ser de vários tipos (SILVA JÚNIOR; BORGES, 2014): a) Temáticos - armazenam conteúdos de assuntos especializados; b) Nacional - agrega a produção acadêmica e/ou científica de um país; c) Institucional – foca na produção intelectual de uma instituição; e d) De pesquisa ou de dados – armazenam dados de pesquisa, sendo esses últimos o foco da pesquisa que originou esse artigo.

Os Repositórios de Dados (RD) surgiram a partir das iniciativas do movimento de acesso aberto, cuja gênese foi a crise dos periódicos científicos na década de 70, mas foi a partir da década de 90 que as publicações científicas eletrônicas fizeram uma mudança radical na Comunidade Científica. Em 2001 foi realizada a reunião de Budapeste que definiu o 1º protocolo de interoperabilidade, o *Open Access Initiative – Protocol for Metadata Harvesting* (OAI – PMH) (BJORK, 2004), constituem hoje uma inovação no sistema de comunicação científica. Nesta reunião também foram definidas duas estratégias para viabilizar o acesso à produção científica: a “via verde” (green road) que implantou os RD institucionais e o autoarquivamento; e a “via dourada” (golden road) que engloba os periódicos científicos eletrônicos de acesso aberto.

De acordo com Sayão; Sales (2016) os RD garantem os princípios de transparência e oferecem um sistema de armazenamento seguro, além da possibilidade de se ter os dados de pesquisa disponíveis on-line, indexados, documentados, para serem acessados, baixados, visualizados e processados por pessoas ou por sistemas, estendendo-os a uma comunidade mais ampla e conectada em rede. De fato, os RD têm sua importância como recurso informacional e se tornam um dispositivo de troca de experiências e compartilhamento de dados científicos¹, além de parte essencial das infraestruturas mundiais de pesquisa em escala global, tornando visível e aberta para toda a sociedade uma parcela importante da atividade de pesquisa, caracterizando a chamada Open Science ou Ciência Aberta² (FORMENTON, 2015).

Sayão; Sales (2016) consideram que os RD, de forma diferente das publicações acadêmicas que falam por si próprias, precisam ter explícitos os seus conteúdos, para poder revelar e transmitir conhecimento no tempo e no espaço, de forma que os dados possam ser interpretados, sintetizados e reanalisados em contextos diversos, com finalidades diferentes das quais foram gerados e coletados originalmente. Assim como é preciso garantir a gestão e preservação a longo prazo, o que faz com que seja necessário o desenvolvimento de metodologias (gerenciais e tecnológicas) que guiem desde a criação dos dados, passando por seu gerenciamento, armazenamento, análise, uso e reuso, em uma diversidade de contextos. O que vem sendo proporcionado pela curadoria de dados, que “engloba atividades de gestão requeridas para manter e gerir dados de pesquisa a longo prazo, de modo que estejam disponíveis para o reuso, favorecendo a colaboração entre pesquisadores e o avanço da ciência” (SIEBRA; BORBA; MIRANDA, 2016, p. 12-13).

¹ Dados científicos são os materiais comumente registrados e aceitos na comunidade científica como necessários para validar os resultados de pesquisa. Ex: fatos e estatísticas recolhidas para posterior referência ou análise, documentos, questionários,

Entre as plataformas disponíveis para criação de um RD encontra-se o Dataverse. Ele é um RD de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa. Pesquisadores, autores de dados, editores, distribuidores de dados e instituições afiliadas recebem o crédito apropriado por meio de uma citação de dados com um identificador persistente (por exemplo, DOI ou Handle) (RICE; SOUTHAL, 2016). Cada conjunto de dados contém metadados descritivos e arquivos de dados (incluindo documentação e código que acompanham os dados) (DATAVERSE PROJECT, 2018, tradução nossa). Essa foi a plataforma utilizada nessa pesquisa.

O Dataverse é um RD de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa. Pesquisadores, autores de dados, editores, distribuidores de dados e instituições afiliadas recebem o crédito apropriado por meio de uma citação de dados com um identificador persistente (por exemplo, DOI ou Handle) (WIKIPEDIA, 2018b). Cada conjunto de dados contém metadados descritivos e arquivos de dados (incluindo documentação e código que acompanham os dados) (DATAVERSE PROJECT, 2018, tradução nossa).

O RD Dataverse fica disponível como um servidor em nuvem e software de código aberto (RICE; SOUTHAL, 2016). O repositório em questão, de acordo com Dataverse Project (2018, tradução nossa), tem se ampliado cada vez mais pela sua visibilidade na WEB, e o crédito acadêmico que proporciona o aumento de contagem das citações com referência a cada dado científico que este repositório custodia. No sítio eletrônico do Projeto Dataverse são caracterizadas quatro categorias de públicos: pesquisadores, artigos, instituições e desenvolvedores em que todas permeiam com objetivo dos dados científicos serem reutilizados pela comunidade científica.

É notório que o Brasil vem sendo afligido por doenças com graves repercussões para a saúde de sua população conforme indicadores operacionais e epidemiológicos, entre elas as Arboviroses (ou Arbovírus), conjunto de patologias formado principalmente pela Zica, Dengue e Chikungunya – de grande poder de morbimortalidade - impacto da doença e de suas mortes em uma sociedade.

Esses três agravos foram sinalizados pelo MS - entidade superior de orientação e formulação de políticas a nível federal que devem ser seguidas pelos profissionais de saúde das esferas Estaduais e Municipais - como uma doença negligenciada, que merece receber atenção especial dos profissionais de saúde que atuam na atenção e na vigilância e dos gestores das esferas federal, estadual e municipal. Segundo a Agência FioCruz de Notícias (2017) doenças negligenciadas são enfermidades tropicais endêmicas, que acomete especialmente as populações pobres da África, Ásia e América Latina. São causadas por agentes infecciosos ou parasitas e são consideradas endêmicas em populações de baixa renda. Essas enfermidades também apresentam indicadores inaceitáveis e investimentos reduzidos em pesquisas, produção de medicamentos e medidas de controle. Elas incapacitam ou matam milhões de pessoas e representam uma necessidade médica importante que permanece não atendida.

Os arbovírus são vírus transmitidos por artrópodes (*Arthropod-borne virus*) e são assim designados não somente pela sua veiculação através de artrópodes, mas, principalmente, pelo fato de parte de seu ciclo reprodutivo ocorrer nesses insetos. São transmitidos aos seres humanos e outros animais pela picada de artrópodes hematófagos, em especial os mosquitos *Aedes aegypti* e *Aedes albopictus* (LOPES; NOZAWA; LINHARES, 2014).

A disseminação de informações sobre o agravo em questão, tanto nos círculos acadêmicos e científicos formais, como nos informais na mídia de massa através das redes sociais, apresenta inúmeras oportunidades de investigação para pesquisadores de várias áreas. É nesta vertente que a proposta do Dataset arboviroses da UFPB está direcionada, contribuindo com dados brutos para o campo científico e para a sociedade diante da urgência da temática. Nessa perspectiva, Freire (2006) em sua pesquisa, ressalta a importância da democratização ao acesso e uso da informação, bem como as vantagens do uso da tecnologia. A informação deve adquirir uma projeção social e econômica na sociedade da informação, com redução de custos.

Ao instalar o *software* Dataverse Network, uma organização pode hospedar vários arquivos virtuais individuais, intitulados *dataverses*, disponibilizando a produção científica, fornecendo uma estrutura de publicação de dados que suporta reconhecimento do autor, citação, descoberta e

transcrições, algoritmos, scripts, etc. (DUDZIAK, 2016).

² Movimento para tornar a pesquisa científica, os dados e a divulgação destes acessíveis a todos os níveis da sociedade (WIKIPÉDIA, 2018).

preservação dos dados. Estes arquivos não exigem custos de *hardware* ou *software*, nem manutenção ou *backups* pelo proprietário dos dados e ainda permitem que toda a visibilidade da *web* e seus créditos sejam transferidos e acessíveis aos usuários. Com esse recurso, os pesquisadores e sua produção científica recebem maior visibilidade e credibilidade dentro da comunidade científica.

Esse artigo teve por objetivo relatar a experiência de criação do 1º RD da UFPB, fazendo uso da plataforma Dataverse, com foco no dataset de arboviroses. Esse RD surgiu após a realização do projeto de pesquisa - Chamada Universal - MCTI/CNPq (Número 01/2016) – intitulado “A Ciência da Informação e a Disseminação de Informações Associadas à Epidemia de Zika Vírus: uma investigação baseada na Análise de Redes Sociais”. E a motivação para o projeto surgiu após o Brasil vivenciar, em 2016, um surto de microcefalia em bebês recém-nascidos, causada pela contaminação de mulheres grávidas com o vírus *zika*. Sendo que para a criação do dataset decidiu-se expandir para um conjunto de patologias formado principalmente pela *Zyca*, Dengue e Chikungunya, denominadas arboviroses. Isso porque essas são doenças com graves repercussões para a saúde da população conforme indicadores operacionais e epidemiológicos.

Fica exposto que este artigo possui como relevância sua originalidade e apresenta o resultado inicial do projeto citado acima. Sua continuidade se dará com os demais autores sociais envolvidos.

2 O ACESSO ABERTO NAS PRODUÇÕES CIENTÍFICAS

O sistema formal de comunicação científica sofreu (e sofre) influências do mundo digital. Os atores sociais – agentes de mudanças – promovem um movimento internacional crescente, afetando as práticas e os modos de produção científica. Além da emergência das tecnologias digitais, os preços abusivos dos periódicos, predispuseram a criação do *Open Access* e suas duas vias: verde (repositório) e dourada (revistas de acesso aberto) (AUTRAN; BORGES, 2014).

Este movimento *Open Access* (Acesso Livre) lutou pela acessibilidade dos documentos na *web*. Surgiu nos anos 90, mas só nos anos de 2000 a 2003 foram realizados três importantes eventos com o objetivo de favorecer este acesso às publicações científicas. Esses eventos foram realizados em Budapeste, Bethesda e Berlim, os quais deram origem a três declarações que, transformaram em relevantes manifestações em favor do acesso aberto às publicações científicas no mundo. Os referidos eventos convocaram autores, associações científicas, editores, universidades e bibliotecas a trabalharem na remoção das barreiras que impedem o acesso livre à literatura científica.

O compartilhamento da informação ocorre quando existem condições adequadas, isto é, o interesse comum, confiança no sentido ético e o uso de padrões mínimos para estabelecer o processo de comunicação. As responsabilidades do pesquisador em disponibilizar os dados brutos de um trabalho para um número maior de pessoas não é menos importante para sua posição, enquanto pesquisador e desenvolvedor de uma pesquisa pioneira. Seu cuidado original na concepção e execução da pesquisa será mantido.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa que originou esse artigo trata-se de uma pesquisa-ação, qualitativa e descritiva (MICHEL, 2009). Na pesquisa-ação os pesquisadores-participantes estão envolvidos de modo cooperativo e participativo na pesquisa, que é concebida e realizada em estreita associação com uma ação ou resolução de um problema coletivo (BALDISSERA, 2001). Na pesquisa qualitativa, segundo Michel (2009), existe a ligação e correlação de dados interpessoais, coparticipação das situações dos informantes e estes são analisados a partir da significação que eles dão aos seus atos. E é descritiva na medida que colhe, interpreta e discute fatos e situações. Além disso, essa é uma pesquisa bibliográfica que fez uso de artigos, livros, teses e dissertações, assim como documental pois fez-se necessário estudar o site e o manual da plataforma Dataverse.

O dataset foco desse relato é um conjunto de dados composto por posts da rede social Twitter sobre as arboviroses, que engloba a *Zyca*, Dengue e Chikungunya. Assim, a coleta de dados para compor o dataset foi realizada no período de outubro de 2017 a março de 2018. Os dados foram coletados por um script feito na linguagem de programação Ruby, fazendo uso de API (Application Programming Interface) disponibilizada pelo próprio Twitter. Os posts coletados foram os que apresentaram uma ou mais das seguintes palavras, desconsiderando maiúsculas e minúsculas: *zika*, *zika*,

zyca, zkv, zikav, dengue, dengue hemorrágica, chikungunya, chicungunya, arbovirose, arboviros, microcefalia.

Os posts identificados, mais de um milhão, foram baixados no formato JasonB e categorizados considerando 3 grupos: Zyca, Dengue e Chikungunya. Em seguida, foi criado um banco de dados modelado para as normas do Dataverse, onde os dados extraídos foram efetivamente armazenados. Ressalta-se que, para poder contextualizar e adicionar valor aos dados, foi criado um conjunto de metadados descritivos para sintetizar os diferentes contextos em que as mensagens foram utilizadas e compreender como os usuários da rede social discutiram/escreveram sobre a temática. O Dataverse foi escolhido como plataforma por ser uma das mais popularmente utilizadas para criação de RDs.

Para realização do trabalho, a equipe foi composta por 6 pessoas: 1 coordenador e 5 bolsistas das áreas de Ciência da Informação e Ciência da Computação.

4 RESULTADOS

No período de outubro de 2017 a março de 2018 foram coletados por um software (*script* feito em *ruby*) utilizando um API (*Application Programming Interface*), disponibilizada pelo próprio Twitter, e baixados no formato JasonB, mais de um milhão de *post* dos usuários da rede social Twitter sobre a temática mencionada acima. Estas mensagens foram categorizadas e em seguida foi criado um conjunto de metadados descritivos para sintetizar os diferentes contextos que as mensagens foram utilizadas e compreender como os usuários da rede social discutiram/descreveram a temática.

O RD da UFPB foi disponibilizado em 2018, no endereço <https://dataverse.ufpb.br/dataverse/root>. A equipe inicial de pesquisadores em uma primeira etapa precisou se familiarizar com o que é e como funciona um repositório de dados, o que foi sanado com a pesquisa bibliográfica. Posteriormente, sobre a plataforma Dataverse. Como toda documentação da plataforma estava em inglês, a equipe se dedicou a traduzir e destacar as principais partes da documentação, antes de começar a implantação da plataforma na instituição. Uma vez a plataforma instalada, ela foi populada com o dataset de arboviroses extraído do Twitter.

Para acessar o RD da UFPB deve-se ir ao sítio descrito acima e com um *login* e senha os pesquisadores poderão acessar o banco de dados. Por padrão, optou-se por permitir acessar o arquivo restrito após o *login* com as credenciais de acesso ao sistema e, durante *download* do arquivo, o sistema deve solicitar a aceitação do termo de acesso.

Para Gomes (2013, p. 29): “[...] a otimização e atualização de dispositivos e ações de informação têm de responder às demandas intensivas de informação e comunicação que satisfaçam, em quantidade e qualidade, às metas e finalidades da promoção da saúde e do atendimento clínico”. O que ressalta as possibilidades de colaboração entre a Ciência da Informação e a área de Saúde. O que pode ser concretizado por exemplo, com a exploração, coleta e organização de bases de dados que possam colaborar com pesquisas na área de saúde, como é o caso do dataset de arboviroses.

Para um melhor gerenciamento de acesso e segurança, buscou-se restringir o acesso aos dados a pesquisadores previamente cadastrados, cujas credenciais de acesso podem ser feitas a partir do cadastro individual do pesquisador no Dataverse, utilizando uma conta do Gmail. Assim, optou-se por permitir acessar o arquivo restrito do dataset apenas após *login* e senha com as credenciais de acesso adequadas ao sistema e, durante *download* do arquivo do dataset, o sistema deve solicitar a aceitação de um termo de acesso e uso. O objetivo é manter um controle de acesso aos dados disponibilizados. Adicionalmente, para aqueles que navegarem nos datasets sem baixar nenhum arquivo há a solicitação de registro no Guestbooks (livro de visitas) de cada dataset. Ressalta-se que o acesso aos dados no RD Dataverse ocorre de quatro formas: aberto, fechado, gerenciado e moderado, o que pode trazer flexibilidade para os casos de dados de pesquisa que necessitem ficarem em sigilo por algum motivo (pesquisa em andamento, patenteamento de solução, etc).

O repositório criado a partir da plataforma Dataverse possui a identificação especificada no Quadro I.

Quadro I – Identificação do Dataverse UFPB

Dataverse	Departamento de CI
Identifier	https://dataverse.ufpb.br/dataverse/dci
Category	Department
Affiliation	CCSA - UFPB
Description	Este repositório concentra dados de pesquisas do Departamento de Ciências da Informação

(DCI) do Centro de Ciências Sociais Aplicadas (CCSA) da Universidade Federal da Paraíba (UFPB).

Fonte: Criado pelas Autoras, 2019.

O campo “Dataverse” descreve o título do banco de dados, o departamento, a universidade ou a revista que conterá os dados. O “Identifier” refere-se a um nome curto a ser utilizado como URL (Uniform Resource Locator - Localizador Padrão de Recursos que é um formato de atribuição universal para designar um recurso na Internet). A “category” identifica o tipo de grupo ao qual a base de dados pertence, no caso, departamento. O campo “Affiliation” deve compor algum nome ou sigla associado à instituição proprietária do repositório, no caso CCSA (Centro de Ciências Sociais Aplicadas) da Universidade Federal da Paraíba (UFPB) e, por fim, “Description” que deve ser a descrição resumida da base de dados, especificando o que é armazenado nela, para exibição na página do repositório.

No Dataverse, os campos de metadados são escolhidos para uso em cada conjunto de dados (dataset) a serem adicionados (THE DATAVERSE PROJECT, 2018, tradução nossa). Considerando a natureza multidisciplinar das pesquisas do Departamento de CI, os pesquisadores optaram pelo uso de campos de metadados gerais, em um padrão que pudesse ser aplicado em dados das diversas áreas do conhecimento. Assim, os principais campos a serem preenchidos no Dataverse para o dataset são: título, autor(es), informações para contato, descrição resumida da pesquisa, data da coleta dos dados no RD, tópicos da pesquisa, palavras chave, entre outros. Esses metadados tanto contextualizam e descrevem os dados armazenados, como facilitam a sua recuperação.

A preservação e acesso a longo prazo são garantidos no Dataverse pela identificação persistente, que protege os documentos digitais com mecanismos que preveem a obsolescência dos dados - migração dos dados para um software mais recente e a prescrição que consiste em guardar o conjunto de bytes para serem consultados quando for necessário. (DATAVERSE PROJECT, 2018, tradução nossa).

5 CONSIDERAÇÕES FINAIS

Nesse cenário, as arboviroses são enfermidades tropicais endêmicas, que merecem receber atenção especial dos profissionais de saúde que atuam na atenção básica e na vigilância em saúde e dos gestores das esferas federal, estadual e municipal (FIOCRUZ, 2017). Elas incapacitam ou matam milhões de pessoas e representam uma necessidade médica importante que permanece não atendida. Assim, o estudo sobre a disseminação de informações sobre essas enfermidades, tanto nos círculos acadêmicos e científicos formais, como nos informais, na mídia de massa através das redes sociais, apresenta inúmeras oportunidades de investigação para pesquisadores. É nesta vertente que a proposta do Dataset arboviroses da UFPB está direcionada, contribuindo com dados brutos para o campo científico e para a sociedade diante da urgência da temática. Podendo-se investigar: o que a população sabe sobre as arboviroses? Que tipo de dúvidas possuem? Como a informação tem chegado até a sociedade? Que tipo de queixas são realizadas sobre as doenças em questão? Que localidades mais discutem sobre arboviroses? Além de ser possível mapear casos de relatos de morte, surtos e agravamento das doenças por meio dos posts coletados.

Ressalta-se que essa pesquisa ainda está no início, porém, a implantação e configuração do dataverse e a coleta dos dados do twitter com a criação do dataset de arboviroses já se mostram como primeiros resultados, que podem trazer diversos desdobramentos no futuro.

Espera-se, como trabalhos futuros, poder relatar os usos feitos do dataset de arboviroses tanto pelos administradores do Dataverse, como por pesquisadores cadastrados na plataforma.

O dataset arboviroses, como parte do RD UFPB/CCSA/Dataverse, possui o intuito de disseminar a informação científica, em especial, sobre as informações informais disseminadas na rede social Twiteer sobre as três doenças. Os dados que este dispositivo custodia, em grande parte, são de grandes pesquisas e devem ser disponibilizados em formato aberto para serem gerados novos conhecimentos e expandir a Comunicação Científica

REFERÊNCIAS

- ABREU, J. P.; VIDOTTI, S. A. B. G. Curadoria Digital nos Contexto dos Repositórios Digitais. In: Encontro Internacional de Dados, Tecnologia e Informação, 2., 2016. Marília. **Anais...** Marília: UNESP, 2016.
- AUTRAN, M. M. M.; BORGES, M. M. Comunicação da Ciência: (r)evolução ou crise?, **RECIIS – Rev. Eletrônica de Comun. Inf. Inov. Saúde**. v. 8, n. 2, p. 122-138, junho, 2014.
- BJORK, B. Open Access to Scientific Publications - an analysis of the barriers to change?. **Information Research**, v. 9, n. 2, January, 2004. Disponível em: <https://helda.helsinki.fi/bitstream/handle/10227/647/bjork.pdf?sequence=1&isAllowed=y>. Acesso em: 02 ago. 2018.
- BALDISSERA, A. Pesquisa-Ação: uma metodologia do conhecer e do “agir” coletivo. **Sociedade em Debate**, Pelotas, v. 7, n. 2, p. 5-25, agosto, 2001. Disponível em: <http://revistas.ucpel.edu.br/index.php/rsd/article/viewFile/570/510>. Acesso em: 02 ago. 2018.
- DATAVERSE PROJECT. **About Dataverse**. Disponível em: <https://dataverse.org/>. Acesso em: 08 jun. 2018.
- FIOCRUZ. **Agência Fiocruz de Notícias**. Disponível em: <https://agencia.fiocruz.br/doen%C3%A7as-negligenciadas>. Acesso em: 08 jun. 2018.
- FORMENTON, D. **Identificação de Padrões de Metadados para Preservação Digital**. São Carlos: UFSCar, 2015. 102 f. Dissertação de Mestrado - Universidade Federal de São Carlos. Disponível em: <https://repositorio.ufscar.br/bitstream/handle/ufscar/7221/DissDF.pdf?sequence=1>. Acesso em: 02 jun. 2018.
- FREIRE, I. M. Acesso à Informação e Identidade Cultural: entre o global e o local. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 58-67, maio/agosto, 2006. Disponível em: <http://www.scielo.br/pdf/ci/v35n2/a07v35n2.pdf>. Acesso em: 02 jun. 2018.
- GOMEZ, M. N. G. O Domínio das Informações em Saúde. In: PINTO, V. B.; CAMPOS, H. H. (Org). **Diálogos Paradigmáticos Sobre Informação para a Área da Saúde**. Fortaleza: Edições UFC, 2013.
- LOPES, N.; NOZAWA, C.; LINHARES, R. E. C. Características Gerais e Epidemiologia dos Arbovírus Emergentes no Brasil. **Revista Pan-Amazônica de Saúde**, v. 5, n. 3, p. 55-64, 2014.
- MICHEL, M. H. **Metodologia e Pesquisa Científica em Ciências Sociais**. 2 ed. São Paulo: Atlas, 2009.
- RICE, R; SOUTHALL, J. **The Data Librarian's Handbook**. London: Facet, 2016. 169p.
- SAYÃO, L. F.; SALES, L. F. Algumas Considerações Sobre os Repositórios de Dados de Pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90 – 115, maio/agosto, 2016. Disponível em: <http://www.uel.br/revistas/informacao/90>. Acesso em: 12 jul. 2018.
- SIEBRA, S. A.; BORBA, V. R.; MIRANDA, M. J. K. F. O. Curadoria Digital: um termo interdisciplinar. In: XVII Encontro Nacional de Pesquisa Em Ciência da Informação (ENANCIB), 17., 2016, Salvador. **Anais...** Salvador, BA: UFBA, 2016. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/4107/2559>. Acesso em: 2 jul. 2018.
- SILVA JUNIOR, L. P.; BORGES, M. M. Preservação Digital no Repositório Científico de Acesso Aberto de Portugal. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 8, n. 4, p. 567-574, out./dez. 2014. Disponível em: <http://www.reciis.icict.fiocruz.br/index.php/receis/article/view/911>. Acesso em: 04 ago. 2018.
- WIKIPÉDIA. **Open Science**. Disponível em: https://en.wikipedia.org/wiki/Open_science. Acesso em: 12 ago. 2018.