

ARTIGO

Recebido em:
20/12/2017

Aceito em:
10/04/2018

Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 23, n. 53, p. 160-173, set./dez., 2018. ISSN 1518-2924. DOI: 10.5007/1518-2924.2018v23n53p160

Plano de Gerenciamento de Dados em Repositórios de Dados de universidades

Data Management Plan in university Data Repositories

Elizabete Cristina de Souza de Aguiar MONTEIRO (beteaguia@yahoo.com.br)*

Ricardo César Gonçalves SANT'ANA (ricardosantana@marilia.unesp.br)**

* Doutoranda em Ciência da Informação pela Universidade Estadual Paulista "Júlio de Mesquita Filho".

** Professor do Programa de Pós-graduação em Ciência da Informação – PPGCI da Universidade Estadual Paulista "Júlio de Mesquita Filho".

Resumo

Plano de Gerenciamento de Dados é o documento formal que descreve os processos de gestão dos dados integrando seu ciclo de vida. A elaboração do Plano de Gerenciamento de Dados auxilia os pesquisadores e os profissionais atuantes nos repositórios. O objetivo deste estudo foi investigar quantos e quais repositórios de dados das 100 melhores universidades do mundo disponibilizam Planos de Gerenciamento de Dados e identificar aspectos relacionados a possíveis benefícios gerados pelo seu uso. A metodologia teve como base a pesquisa documental e exploratória de natureza qualitativa e quantitativa. Foi utilizada a metodologia exploratória para fazer o levantamento das universidades através do sítio *webometrics.info* e dos respectivos repositórios. Os resultados demonstram que apenas 36 repositórios disponibilizam Planos de Gerenciamento de dados e que suas instruções variam dependendo das características dos repositórios e dos conjuntos de dados neles depositados. Concluiu-se que seu uso proporciona diversos benefícios a todos os atores envolvidos na gestão de dados, como a indicação que os responsáveis pelos dados devem ser citados destacando a valorização do trabalho da coleta, a maximização da reutilização dos dados, as indicações de quais licenças estão atribuídas aos conjuntos de dados, o que determina como os dados são licenciados e as formas de utilização e a orientação sobre aspectos de privacidade dos dados.

Palavras-chave: Plano de Gerenciamento de Dados; Gestão de dados; Dados científicos; Repositório de dados. Ciclo de Vida dos Dados.

Abstract

Data Management Plan is the formal document that describes the data management processes integrating their life cycle. The elaboration of the Data Management Plan assists the researchers and professionals working in the repositories. The objective of this study was to investigate how many and which data repositories of the world's 100 best universities provide Data Management Plans and identify aspects related to possible benefits generated by their use. The methodology was based on documental and exploratory research of a qualitative and quantitative nature. The exploratory methodology was used to survey the universities through the *webometrics.info* website and its repositories. The results show that only 36 repositories make available Data Management Plans and that their instructions vary depending on the characteristics of the repositories and the data sets deposited in them. It is concluded that its use provides several benefits to all actors involved in data management such as the indication that the data controllers should be mentioned, highlighting the value of the collection work, the maximization of data reuse, the indication of which licenses are assigned to the datasets, which determines how the data is licensed and the forms of data collection, use and guidance on data privacy issues.

Keywords: Data Management Plan; Data management; Search data; Data repository. Data Life Cycle.



v. 23, n. 53, 2018.
p. 160-173
ISSN 1518-2924



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

1 INTRODUÇÃO

Dados científicos fazem parte do cotidiano do processo de pesquisa. A necessidade de acesso e compartilhamento de dados é reconhecida, está evidente nos documentos de planejamento de muitos projetos e nas colaborações científicas internacionais e a reutilização dos dados proporciona benefícios à comunidade científica (FITZGERALD; FITZGERALD; PAPPALARDO, 2011; PINFIELD; COX; SMITH, 2014).

Assim, as instituições acadêmicas e científicas passaram a ter cada vez mais a responsabilidade no gerenciamento de dados científicos coletados ou produzidos em grande quantidade, velocidade e variedade por pesquisadores nas diversas áreas do conhecimento.

Esse crescente volume de dados criou desafios em sua gestão ao longo de seu ciclo de vida. A gestão apropriada de dados é um pré-requisito para o seu compartilhamento eficaz entre uma comunidade científica específica e comunidades que não fazem parte do grupo alvo, contudo podem se beneficiar desses dados, principalmente, dados de observação única, que só tem a possibilidade de coletá-los uma única vez.

A gestão de dados requer, por parte de seus detentores, planejamento e ações concretas que tragam eficiência não só para coleta e armazenamento como também e, principalmente, para fase de recuperação desses dados ampliando sua visibilidade e potencial uso (SANT'ANA, 2016).

Na ambiência da investigação científica, esse processo que agrega valor configura-se como importante fator para a ampliação do potencial de impacto dos resultados das pesquisas e da própria instituição (PRYOR, 2012). Itens documentários que publicam os dados que foram usados para embasar seus resultados recebem maior número de citações (PIWOWAR; VISION, 2013).

Desse contexto emerge a necessidade de políticas para a gestão dos dados envolvidos nas pesquisas. Agências de fomento como *National Science Foundation* (NSF), *National Institutes of Health* (NIH), *National Oceanographic Data Center* (NODC) e NASA dos Estados Unidos, Horizon2020 da Europa, AHRC, BBSrc, *Cancer Research UK*, EPSRC, ESRC, MRC, NERC, STFC, *WELLCOMETrust* no Reino Unido estão incentivando, orientando ou mesmo tornando obrigatória a elaboração de Plano de Gerenciamento de Dados (PGD) para os projetos que terão o financiamento de suas pesquisas por essas agências (CORRÊA COUTO, 2016).

No Brasil, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) lançou a *Chamada de Proposta de Pesquisa - 2017* em seu *Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais (PFPMCG)* que exige junto a outros documentos o PGD para o pesquisador que solicitar o financiamento (FAPESP, 2017a). A obrigatoriedade se estenderá gradativamente, a partir de 2018, para outras modalidades de apoio, como as propostas de *Auxílio à Pesquisa – Regular, Jovem Pesquisador, bolsas de Doutorado e de Pós-doutorado* (PLANOS..., 2017). As universidades do Estado de São Paulo terão que implementar repositórios de dados, ou incluir os conjuntos de dados nas coleções de seus Repositórios Institucionais, como a Universidade Estadual Paulista (UNESP), para atender sua comunidade no ciclo de vida dos dados (CVD).

Os Repositórios de dados científicos são ambientes digitais implementados nas universidades com infraestrutura para dar suporte aos pesquisadores na gestão e na disponibilização de dados científicos o que potencializa a reutilização por outros pesquisadores (MONTEIRO, 2017).

As universidades internacionais estão implementando repositórios de dados para armazenar e disponibilizar os dados científicos da sua comunidade. Os repositórios de dados disponibilizam em seus sítios o Plano de Gerenciamento de Dados (PGD) com diretrizes para orientar os pesquisadores sobre o processo de depósito de seus dados, os profissionais que atuam no repositório e a comunidade que fará uso do repositório para a busca e coleta de dados. As universidades brasileiras tendem a seguir na mesma direção (FAPESP, 2017b).

O PGD é um documento elaborado no início do projeto de pesquisa no qual são descritas as diretrizes para o ciclo de vida dos dados. Assim, os procedimentos descritos para a gestão de dados são documentados no PGD, descrições essas que são direcionadas àqueles que estão envolvidos de alguma forma com a gestão desses dados (MONTEIRO, 2017). Nesse processo de descrição dos repositórios, devem-se observar suas políticas, os objetivos e as características da comunidade atendida.

O objetivo deste artigo é apresentar os resultados de uma investigação sobre repositórios de dados das 100 melhores universidades do mundo que verificou quantos e quais disponibilizam em seus sítios Planos de Gerenciamento de Dados e identificou aspectos relacionados a possíveis benefícios gerados pela adoção destes PGDs.

A coleta, armazenamento, recuperação e descarte de dados nos repositórios de dados, associados aos aparatos tecnológicos, em resumo, constituem alguns dos elementos decisivos para o estudo das relações entre dados, tecnologia e sociedade. Sendo assim, este artigo tem o intuito de contribuir com todos aqueles que de alguma forma estão envolvidos com as fases do ciclo de vida dos dados de repositórios de dados discutindo os benefícios do uso do Plano de Gerenciamento de Dados.

2 REPOSITÓRIOS DE DADOS

A gestão de dados científicos vem ganhando atenção crescente nos últimos anos, devido, em grande parte, à conscientização sobre o valor contido nos dados científicos e dos riscos de perder esses dados ao longo do tempo (SESARTIC; TOWE, 2016). A inevitabilidade de infraestrutura para o planejamento, implementação e gestão alterou a configuração de serviços necessários ao contexto de repositórios digitais.

Os repositórios digitais começam a ser implantados em vários países e, em sua maioria, estão vinculados a universidades ou instituições de pesquisa (OPENDOAR, 2017). Parte desses repositórios é de acesso aberto, sendo 3.448 no total, desses, 14,5% estão nos Estados Unidos, 26,6% estão em países da Europa e 2,8% no Brasil (OPENDOAR, 2017). Nota-se que a grande maioria, 41,1%, está em países do continente Norte-Americano e Europeu.

Os repositórios de dados têm sua gênese com a necessidade de gestão dos dados científicos, estão vinculados às universidades e instituições de pesquisa e contribuem para assegurar que os dados sejam publicados e disponibilizados para a comunidade científica com o menor número possível de restrições. Esses repositórios estão implementados em vários países, são mais de 1.100 com a grande maioria nos Estados Unidos e Europa (KING; FORCE, c2017).

Conforme Pinfield (2009, p. 165, grifo nosso, tradução nossa)

Um repositório pode ser definido como um conjunto de sistemas e serviços que facilita o armazenamento, gerenciamento, recuperação, exibição e reutilização de objetos digitais. Os repositórios podem ser estabelecidos por instituições, comunidades, financiadores de pesquisa ou outros grupos. Eles podem fornecer acesso a uma variedade de objetos digitais, incluindo artigos de revistas revisados por pares, capítulos de livros, teses, conjuntos de dados, objetos de aprendizado ou arquivos *rich media*.

Esses repositórios contribuem na gestão de grandes quantidades de dados. Os repositórios de dados são mantidos por conjuntos de ações que viabilizem o armazenamento de dados visando à otimização da coleta pelos pesquisadores, o que amplia as potencialidades de reuso destes dados (MONTEIRO, 2017).

Desafios no âmbito da Ciência da Computação e da Ciência da Informação, tais como àqueles que ocorrem em todas as fases do CVD, Preservação, Disseminação, Direitos Autorais, Qualidade, Integração e Privacidade, permanecem em aberto o que torna difícil descobrir, compartilhar ou reutilizar dados, pois:

1. dados valiosos podem ter sido descartados;
2. tecnologias da informação tendem a ter processo de obsolescência altamente acelerado;
3. formatos incompatíveis podem tornar os dados difíceis ou impossíveis de integrar;
4. o fluxo de dados entre domínios pode ser impedido por metadados incompletos, imprecisos e/ou mal descritos;
5. muitos cientistas relutam em compartilhar dados devido à falta de recompensa, às questões de propriedade intelectual e documentação apropriada (LEE et al., 2009).

Geralmente a instituição que implementa repositório de dados tem em seu sítio um documento ou informações relacionadas ao Ciclo de Vida dos Dados e o Plano de Gerenciamento de Dados, para orientar os pesquisadores que depositarão conjuntos de

dados, os que utilizarão os repositórios como fonte para coleta de dados e os funcionários que atuam no repositório. Cada fase e fator do ciclo de vida dos dados devem ser considerados na elaboração do PGD que vão contribuir para diminuir os desafios elencados por Lee et al. (2009).

3 PLANO DE GERENCIAMENTO DE DADOS

Gestão de dados está gradativamente despontando como protagonista no processo de investigação científica devido às exigências de agências de financiamento, instituições de pesquisa e líderes políticos que liberam recursos para pesquisa além de revistas científicas para liberação e publicação de dados científicos no momento da publicação do artigo (WALLIS; ROLANDO; BORGMAN, 2013; BORGMAN et al., 2015).

O gerenciamento de dados envolve o processo em que as observações, medidas e análises são definidas, feitas e documentadas e os métodos pelos quais os dados dessas observações são posteriormente processados e mantidos (WILLIAMS; BAGWELL; ZOZUS, 2017). A gestão de dados é uma parte essencial de todos os esforços de pesquisa e viabiliza a replicação e a reprodução dos dados, a avaliação das afirmações científicas, a verificação dos resultados publicados e a realização de análises alternativas. A replicação de dados, comum devido ao compartilhamento interpessoal, é potencializado com a publicação dos dados em repositórios de dados.

Os requisitos para depósito em repositórios de dados, a publicação de dados em periódicos científicos junto ao artigo como nas revistas *Nature*, *Plos One* e *Science*, a publicação de dados em periódicos de dados como o *Biodiversity Data Journal*, o *Data in Brief da Elsevier*, o *Earth system Science Data*, o *GeoScience Data Journal*, o *Journal of Open Archeology Data*, o *Journal of Open Psychology*, o *Journal of Physical and chemical Research Data*, o *Journal of Open Research Software* e o *Open Health Data*, as políticas de compartilhamento de dados e o Plano de Gerenciamento de Dados são realidades na comunidade científica (MONTEIRO, 2017, WALLIS; ROLANDO; BORGMAN, 2013).

Um objetivo de destaque na gestão de dados científicos é assegurar que os mesmos possam ser compreendidos e interpretados por outros pesquisadores ao longo do tempo. Para isso é essencial uma descrição clara e detalhada dos dados, anotações adicionais e informações que contextualizam os dados e possibilitem que transmitam informação e conhecimento no tempo e no espaço (SAYÃO; SALES, 2015).

Para os documentos, as informações e instruções referentes ao gerenciamento dos dados científicos disponibilizados nos repositórios foi adotado, neste artigo, o termo Plano de Gerenciamento de Dados para padronização, sendo que os sítios analisados denominaram com mais de uma forma como *Data Management Planning*, *Research Data Management*, *Data Management* e *Policy and Terms on Use*.

O Plano de Gerenciamento de Dados é um plano que descreve diferentes atividades e processos associados ao ciclo de vida dos dados e envolve “[...] a concepção e criação de dados, armazenamento, segurança, preservação, recuperação, partilha e reutilização, todos tendo em conta as capacidades técnicas, considerações éticas, questões legais e estruturas de governança.” (COX; PINFIELD, 2014, tradução nossa).

Os PGDs dos repositórios relacionam e descrevem os aspectos e as diretrizes que envolvem o gerenciamento dos dados científicos, os quais estão em consonância com as exigências das agências de fomento. Os PGDs não seguem um modelo padronizado, cada um segue um tipo de estrutura e variam entre fornecer as informações aos pesquisadores para a elaboração de seus PGDs e informações da gestão dos dados no repositório.

Os procedimentos adotados na execução de um PGD definem e estabelecem métodos de execução das atividades e detalham os procedimentos que serão realizados. O planejamento é um processo cíclico, dinâmico e interativo, em que as fases não precisam ser lineares, pois há uma dinâmica no processo (ALMEIDA, 2005).

Conjuntos de dados de um determinado grupo de pesquisadores podem conter diferentes formatos, tipos e descrições, tornando-os altamente heterogêneos e, à medida que o tamanho dos conjuntos de dados aumenta, o seu gerenciamento tende a se tornar árduo (LEE et al., 2009).

O PGD auxilia tanto os pesquisadores que coletam e manipulam conjuntos de dados quanto àqueles profissionais que atuam nos repositórios de dados científicos e fornece diretrizes para todo o ciclo de vida dos dados, com destaque para a indicação dos tipos e

formatos dos dados, os métodos de compartilhamento de dados e as políticas para reutilização e redistribuição de dados (COUTO CORRÊA, 2016; VAN LOOAN et al., 2017; WILLIAMS; BAGWELL; ZOZUS, 2017).

Uma análise de qualidade feita em 119 PGDs enviados à *National Science Foundation* (NSF) com solicitações de financiamentos pelos pesquisadores da *Wayne State University* fornece indícios dos elementos essenciais que constam nos PGDs dos repositórios seguindo os requisitos da NSF:

- Elementos básicos: responsabilidade individual, quantidade de dados, tipos e formatos esperados, descrição/metadados, *backup* dos dados, intenção de compartilhamento dos dados, duração da preservação dos dados;
- Método de compartilhamento de dados: e-mail, *website* pessoal ou base de dados, artigos ou conferências, material suplementar, repositório de dados;
- Políticas de compartilhamento de dados: reuso ou distribuição, informação sensível, propriedade intelectual (VAN LOOAN et al. 2017).

Para contextualizar a coleta de dados, este artigo utilizou o Ciclo de Vida dos Dados (CVD), modelo composto por quatro fases: Coleta, Armazenamento, Recuperação e Descarte, sobre as quais perpassam por seis fatores como: Preservação, Disseminação, Direitos Autorais, Qualidade, Integração e Privacidade (SANT'ANA, 2016).

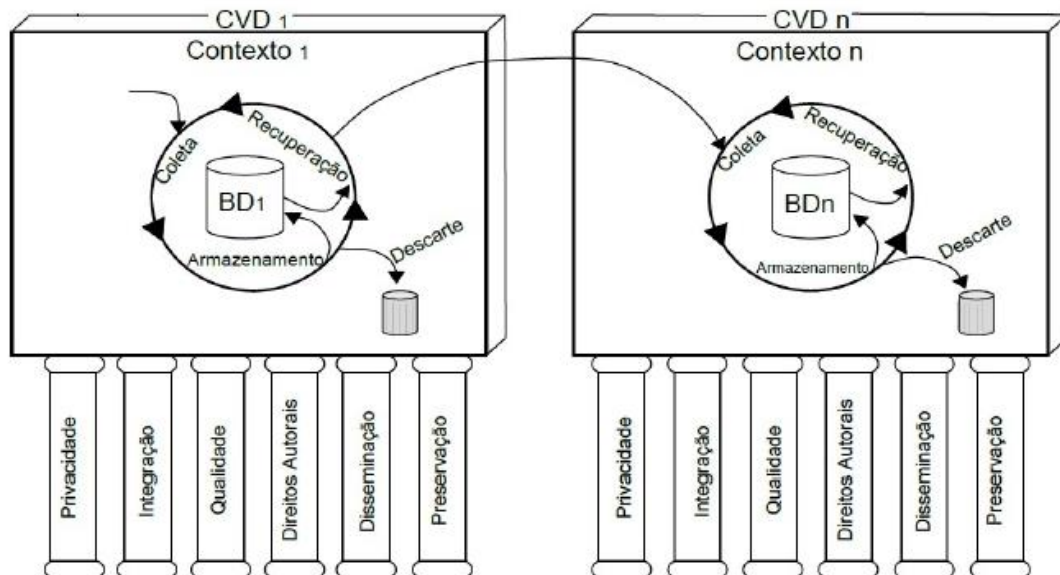


Figura 1: Ciclo de Vida dos Dados para a Ciência da Informação (CVD-CI)
Fonte: SANT'ANA, 2016

A fase da coleta caracteriza o processo de obtenção dos dados em que têm as atividades “[...] vinculadas à definição inicial dos dados a serem utilizados, seja na elaboração do planejamento de como serão obtidos, filtrados e organizados, identificando-se a estrutura, formato e meios de descrição que será utilizado.” (SANT'ANA, 2013, p. 18).

Diversos atores participam no contexto do acesso a dados científicos, entre eles: o pesquisador 1 que armazena seus dados no repositório, o Staff (profissional responsável pelos dados no repositório) e a sociedade (pesquisadores 2, 3 e 4 que farão coleta nos repositórios). Todos os participantes utilizam PGDs para a gestão dos dados. O PGD do repositório pode fornecer subsídios para o pesquisador 1 desenvolver seu PGD e para os pesquisadores P2, P3, P4 na coleta e uso dos dados publicados pelo repositório (FIGURA 2).

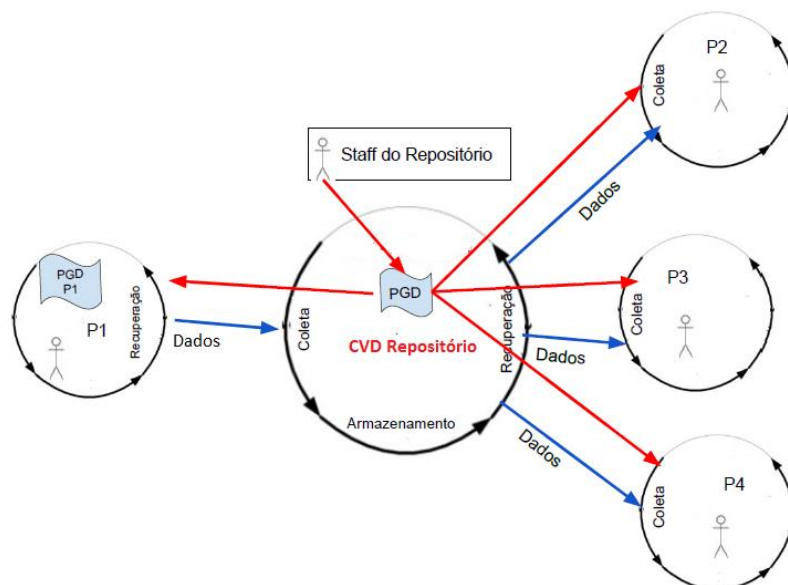


Figura 2: Ciclo de Vida dos Dados no Repositório

Fonte: Elaborada pelos autores.

Observa-se na figura 2 que o PGD do repositório respalda todos os atores envolvidos com os dados dos repositórios:

- **o Pesquisador 1:** o pesquisador descreve em seu PGD como os dados serão coletados e em qual repositório os dados serão armazenados, estabelecendo critérios relacionados às questões legais referentes ao direito autoral, como os dados poderão ser utilizados já empregando uma licença de uso e as questões de privacidade dos sujeitos participantes da pesquisa, caso a pesquisa tenha sujeitos humanos que possam ser identificados, tudo de acordo com as especificações da agência que financiou sua pesquisa ou do repositório que vai armazenar os dados;
- **o CVD do repositório:** o Staff desenvolve o PGD de acordo com as características, missão e objetivos do repositório. O PGD fica disponível para o P1 que vai depositar os dados, para os P2, P3 e P4 que farão busca e coleta de dados no repositório e para os profissionais que atuam no repositório;
- **os Pesquisadores 2, 3 e 4:** realizam suas coletas com base nas funcionalidades oferecidas pela fase de recuperação do repositório e têm no PGD informações sobre todo o CVD dos dados arquivados. Essas informações orientam os pesquisadores sobre como poderão utilizar os dados.

Todos os atores envolvidos no gerenciamento de dados se beneficiam com o uso de Plano de Gerenciamento de Dados. Os repositórios fornecem orientações para a comunidade sobre os aspectos de compartilhamento e uso legal e ético dos dados.

4 PROCEDIMENTOS METODOLÓGICOS

A metodologia teve como base a pesquisa documental e exploratória de natureza qualitativa e quantitativa. A coleta de dados se iniciou com a busca das melhores universidades do mundo por meio do ranking *webometrics.info* definindo o escopo com as 100 melhores ranqueadas. A localização dos repositórios de dados nas universidades foi realizada nos meses de julho a setembro de 2016. Em seguida foi realizada a pesquisa exploratória para o levantamento das páginas oficiais das universidades identificadas para localização dos repositórios de dados e de seus PGDs para a identificação dos benefícios de sua utilização. Toda a coleta foi realizada através dos sítios localizados. Para a localização dos PGDs foram averiguados todos os *links* do sítio oficial dos repositórios de dados observando informações que abordassem sobre Plano de Gerenciamento de Dados, políticas de uso, gerenciamento de dados e dados e informações no item Sobre o repositório de dados. O processo de recuperação dos dados foi realizado por meio de coleta dos Planos de Gerenciamento de Dados dos repositórios de dados encontrados.

5 RESULTADOS E DISCUSSÃO

A análise incluiu a identificação dos repositórios de dados das universidades e dos seus Planos de Gerenciamento de Dados.

A figura 3 ilustra os caminhos que foram seguidos para a coleta nos repositórios de dados. Os resultados indicam que das 100 universidades analisadas, 55 dispõem de repositórios de dados. Desses, 36 têm PGDs, os quais foram analisados.

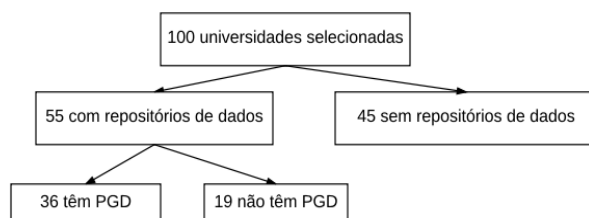


Figura 3: Direcionamento das análises

Fonte: MONTEIRO; SANT'ANA, 2017.

As universidades que têm repositórios e que disponibilizam PGDs estão apresentadas no quadro 1. Na primeira coluna está indicando a ordem em que aparecem no quadro, na segunda coluna o país, na terceira coluna o nome da Universidade, na quarta coluna está o nome do repositório, na quinta coluna o endereço eletrônico do repositório e na sexta coluna a ferramenta para elaboração do PGD que o repositório indica, quando localizada no repositório.

Nº	PAÍS	UNIVERSIDADE	REPOSITÓRIO	SITIO	FERRAMENTA
1	EUA	Harvard University	Harvard Dataverse	https://dataverse.harvard.edu/	
2	EUA	Massachusetts Institute of Technology	DSpace @ MIT	http://dspace.mit.edu/	DMPTool
3	EUA	Stanford University	Stanford Data Repository	https://sdr.stanford.edu/	DMPTool
4	EUA	University of California Berkeley	Data Repository	http://mwhite.berkeley.edu/	DPMTTool
5	EUA	University of Michigan	Deep Blue Data	https://deepblue.lib.umich.edu/data/?locale=en	
6	EUA	University of Washington	ResearchWorks Archive	http://digital.lib.washington.edu/index.html	DMPTool
7	EUA	University of Pennsylvania	ScholarlyCommons	https://repository.upenn.edu/	DMPTool
8	UK	University of Oxford	Oxford University Research Archive	http://researchdata.ox.ac.uk/	DMPOnline
9	EUA	University of California Los Angeles UCLA	Merritt	https://merritt.cdlib.org/	DMPTool
10	EUA	Yale University	ISPS Data Archuve	https://isps.yale.edu/research/data	
11	UK	University of Cambridge	Data Repository	https://www.data.cam.ac.uk/repository	DMPOnline
12	EUA	University of Wisconsin Madison	Research Data Service	http://researchdata.wisc.edu	
13	EUA	Michigan State University	RDP MSU Code Repository	http://rdp.cme.msu.edu/ https://tech.msu.edu/technology/website-services/code-repository/	
14	EUA	University of Texas Austin	Texas ERC	https://texaserc.utexas.edu/erc-data/data-type/	
15	EUA	University of California San Diego	iDASH	https://idash.ucsd.edu/data-repository-0	
16	EUA	Pennsylvania State University	ScholarShere	https://scholarsphere.psu.edu/catalog?f%5Bresource_type_sim%5D%5B%5D=Dataset&q=	DMPTool

17	EUA	University of Illinois Urbana Champaign	Illinois Data Bank IDEALS	https://databank.illinois.edu/ https://www.library.illinois.edu/scp/repositories/	
18	EUA	University of North Carolina Chapel Hill	Carolina Digital Repository	https://dataverse.unc.edu/ https://cdr.lib.unc.edu/	
19	EUA	Princeton University	DataSpace	http://dataspace.princeton.edu/jspui/	
20	UK	University College London	PARNASSUS	http://www.ucl.ac.uk/parnassus/partners_area/data_repository	DMPOnline
21	CA	University of British Columbia	ABACUS cIRcle	http://dvn.library.ubc.ca/dvn/ https://circle.ubc.ca/about/	
22	EUA	University of Maryland Baltimore	GIS DRUM	https://one.umd.edu/task/resources/gisdata https://drum.lib.umd.edu/	
23	EUA	Purdue University	Purdue University Research Repository	https://purr.purdue.edu/	DMPTool
24	EUA	California Institute of Technology Caltech	Caltech	https://data.caltech.edu/	
25	EUA	University of Virginia	Libra Data	https://dataverse.lib.virginia.edu/	
26	EUA	University of California Irvine	UC Irvine Machine Learning Repository	http://archive.ics.uci.edu/ml/index.php	
27	EUA	University of Arizona	UA Campus Repository	http://arizona.openrepository.com/arizona/	DMPTool
28	UK	University of Edinburgh	Edinburgh DataShare	http://datashare.is.ed.ac.uk/	
39	EUA	Washington University Saint Louis	CNDA	http://nrg.wustl.edu/nrg-projects/cnda/	DMPTool
30	CAN	Simon Fraser University	Radar	https://researchdata.sfu.ca	
31	NDL	Utrecht University / Universiteit Utrecht	Utrecht University	https://uu.figshare.com/	
32		Virginia Polytechnic Institute and State University	VTechData	https://data.lib.vt.edu/	
33	EUA	Tufts University	GIS	http://sites.tufts.edu/gis/data-sources/data-source-index/haiti-geospatial-data-resources/	DMPTool
34	GER	Ruprecht Karls Universität Heidelberg	heiDATA	https://heidata.uni-heidelberg.de/	DMPOnline
35	DK	University of Copenhagen/ Københavns Universitet	GBIF	http://danbif.au.dk/ipt/resource?r=ds17	
36	NDL	University of Amsterdam/ Universiteit van Amsterdam	UvA / AUAS figshare	http://rdm.uva.nl/en/support/uva-auas-figshare/introduction.html	DMPOnline

Quadro 1: Repositórios das universidades

Fonte: Elaborado pelos autores.

O gráfico 1 apresenta a quantidade de repositórios de dados nas universidades de cada país com base nas 100 melhores universidades analisadas.

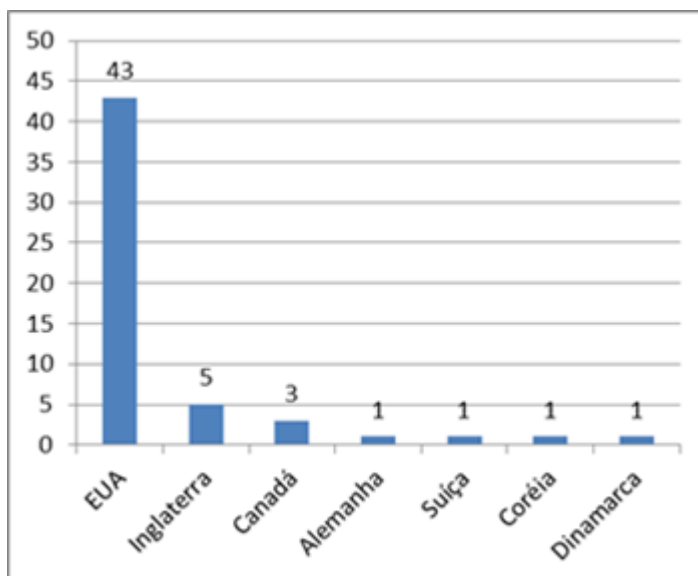


Gráfico 1: Quantidade de Repositório de Dados por país

Fonte: MONTEIRO, 2017.

Nota: Baseado nas cem melhores universidades do mundo

Observa-se no gráfico 1 que a maior quantidade dos repositórios de dados que foram localizados está nas universidades dos Estados Unidos no total de 43, seguido da Inglaterra com cinco e Canadá com três. Os países com menor quantidade de universidades com repositórios de dados são a Alemanha, a Suíça, a Coreia e a Dinamarca com um em cada universidade. Evidencia-se que prevalecem repositórios nas universidades de idioma inglês.

O gráfico 2 ilustra a quantidade de repositórios de dados que disponibilizam PGDs em seus sítios. Nota-se que, dos 55 repositórios de dados das universidades apresentadas, 36 deles dispõem de PGDs, correspondendo a 65% do total e, em 19 deles, não foram localizados PGDs correspondendo a 35% da totalidade.

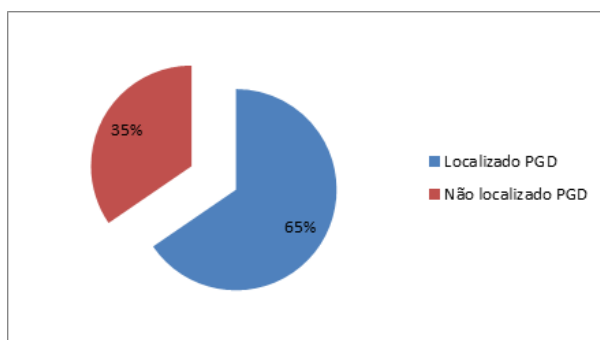


Gráfico 2: Relação de Repositórios de Dados com PGDs

Fonte: MONTEIRO, 2017.

Os repositórios de dados documentam as instruções e normativas nos PGDs nos quais mencionam os vários aspectos do ciclo de vida dos dados. As instruções inclusas nos PGDs variam dependendo das características dos repositórios e dos conjuntos de dados neles depositados. Percebe-se que cada repositório elaborou seu PGD de acordo com as necessidades e particularidades de sua comunidade e do tipo de conteúdo abordado nos conjuntos de dados.

Destaca-se que os repositórios indicam o DMPTool e o DMP Online, duas ferramentas para a elaboração de PGD incluindo os requisitos necessários para tal. Conforme o quadro 1, o DMPTool foi indicado por 11 repositórios e o DMP Online foi indicado por cinco repositórios.

O DMPTool¹ que é uma ferramenta da Universidade da Califórnia e que fornece orientações sobre instituições financiadoras específicas que exigem PGD e um guia para a elaboração do documento. Os PGDs tem sua propriedade intelectual vinculada a quem os criou. O pesquisador que elabora o PGD no DMPTool pode optar em compartilhar seu PGD publicamente contribuindo com outros pesquisadores (DMPToll, 2017).

Os usuários do DMPTool podem visualizar amostras de PGDs, requisitos das agências financiadoras e exibir as alterações mais recentes feitas em seus planos uma vez que a ferramenta permite ao usuário criar um documento editável para apresentar a uma agência de financiamento. Pode-se, ainda, acomodar versões diferentes à medida que os requisitos de financiamento mudam (DMPTool, 2017).

A ferramenta DMPTonline² foi desenvolvida pelo *Digital Curation Center* (DCC) do Reino Unido, com a colaboração do *Curation Center* da Universidade da Califórnia (UC3) com intuito de ajudar os usuários criar, analisar e compartilhar PGDs (DMPTonline, 2017). O DMPTonline fornece orientações personalizadas, exemplos e modelos para ajudar os pesquisadores no desenvolvimento de seus PGDs e atender aos requisitos das agências financiadoras do Reino Unido e de outros países (DMPTonline, 2017).

A elaboração e a disponibilização de PGD por repositórios de dados podem proporcionar benefícios a todos os atores envolvidos (Figura 2) na gestão dos dados, pois orienta sobre os vários aspectos conforme descrito nos repositórios analisados.

No quadro 2 são apresentados os aspectos e os respectivos benefícios que o uso do PGD promove. Os aspectos estão relacionados às instruções descritas nos PGDs para orientar os pesquisadores no desenvolvimento dos planos para gestão de seus dados ou aspectos relacionadas às atividades desenvolvidas e oferecidas pelos repositórios. Na primeira coluna estão relacionados os aspectos abordados, na segunda coluna estão relacionadas os benefícios e na terceira coluna estão identificados os repositórios que foram localizados os aspectos. Os números dos repositórios são a ordem elencadas no quadro 1.

ASPECTOS	DESCRIÇÃO DOS BENEFÍCIOS	Repositório
Formato	Indicam quais são os formatos dos dados que são armazenados no repositório. Orientam sobre quais formatos de arquivo têm maior probabilidade de serem acessíveis no futuro por serem proprietários ou livre.	1, 4, 5, 7, 8, 17, 19, 24, 28, 29, 30, 3, 34, 361
Integridade	Assegura que o repositório manterá a integridade dos dados	1, 17
Proveniência	Indica a importância ou a necessidade de descrever a proveniência dos dados	17
Tamanho	Os PGDs não são de tamanhos únicos, podem variar de acordo com as especificações orientadas pelas agências de fomento ou pelas diretrizes dos repositórios	todos os PGDs
URL	Os dados recebem um URL no repositório	3, 5, 17, 31
Identificador persistente	Os dados recebem um identificador persistente como o identificador de objeto digital (DOI) para que o pesquisador possa conectar seus dados para suas publicações. O repositório traz orientações sobre a utilização do DOI	1, 5, 8, 11, 17, 20, 25, 29, 32, 34, 36
Acesso aos dados	Opções de acesso flexível em que os dados ficam acessíveis a todos, ou com acesso restrito mediante cadastro ou solicitação, dependendo das opções do pesquisador	1, 18
Citação	Exemplos de como citar os conjuntos de dados padronizando a citação, fator que indica aspectos relacionados aos direitos autorais	1,9, 23, 26, 27, 29, 30, 36

1 <https://dmptool.org/>

2 <http://www.dcc.ac.uk/dmponline>

ASPECTOS	DESCRIÇÃO DOS BENEFÍCIOS	Repositório
Valorização do trabalho da coleta	Indicação que os responsáveis pelos dados devem ser citados destacando a valorização do trabalho da coleta	1, 2, 9, 10, 11, 17, 23, 27, 29
Referência	Indicação de como fazer a referência dos dados nos modelos de normalização de referências	1
Metadados	Orienta sobre o uso de metadados, pois seu uso fornece descrição detalhada o que impacta a descoberta dos dados	1, 3, 4, 8, 9, 17, 19, 23, 24, 27, 29, 30
PGD - elaboração	Indicações de ferramentas que auxiliam os autores a montarem seus PGDs	2, 3, 4, 6, 7, 8, 9, 11, 16, 20, 23, 27, 29, 33, 36
PGD	Indicação das agências que solicitam o PGD e exemplos de PGDs submetido às agências	1, 3, 4, 8, 9, 16, 23, 24, 27, 30
Licença de uso	Indicação de quais licenças estão atribuídas aos conjuntos de dados, o que determina como os dados são licenciados e as formas de utilização	3, 4, 5, 8, 9, 10, 11, 17, 23, 25, 27, 28, 29, 30, 31, 32, 34, 35, 36
Agências de fomento	Auxiliar ao pesquisador que vai depositar os dados a atender aos requisitos das agências de fomento que se aplicam aos seus conjuntos de dados	1, 2, 3, 4, 7, 9, 11, 20, 25, 27, 29, 30, 34
Privacidade	Orienta sobre aspectos de privacidade dos dados para manter a privacidade dos sujeitos referenciados	1, 3, 4, 5, 7, 8, 9, 11, 17, 19, 20, 25, 27, 28, 29, 30, 31, 36
Lei	Indicação de qual lei ou normas de instituição que são aplicados para manter a privacidade dos sujeitos referenciados	3, 4, 5, 8, 25, 27, 32
Embargo	Informações sobre embargo dos dados	1, 17, 18, 19, 28, 36
<i>Backup</i>	Armazenamento, backup, replicação, controle de versão e descarte em que indica se os arquivos dos dados são copiados regularmente, se existem réplicas em locais diferentes e se as versões antigas dos dados são mantidas ou descartadas	1, 5, 17, 19, 20, 24, 25, 30, 34, 36
Custos	Indicação se há custos para a preparação dos dados, de armazenamento ou para acesso a longo prazo	1, 5, 8, 11, 12, 20, 30, 31, 34
Segurança do sistema	Requisitos de segurança do sistema (tempos limite de sessão ociosa, desativação de contas genéricas, inibição de adivinhação de senha), requisitos operacionais (relatórios de violação, <i>patches</i> , complexidade de senha, registro) e auditoria e revisão regulares	1, 18
Segurança dos dados	Orienta sobre como manter a segurança dos dados (como exemplo a criptografia)	4, 5, 7, 18, 24, 34, 36
Preservação/curadoria	Serviço para a preservação dos dados ou orientações de como preservar os dados	1, 3, 5, 8, 12, 17, 18, 19, 32

Quadro 2: Benefícios do uso de PGD

Fonte: Elaborado pelos autores

Observa-se no quadro 2 que há aspectos que são apresentados e descritos por um maior número de repositórios como formato, metadados, licenças de uso, agências de fomento e privacidade.

Os PGDs dos repositórios:

- englobam todo o ciclo de vida dos dados;
- disponibilizam os protocolos de tratamento dos dados, o plano de segurança e de compartilhamento dos dados, os manuais de procedimentos de operações;
- fornecem diretrizes para a gestão dos dados para os profissionais atuantes nos repositórios;
- refletem as necessidades dos pesquisadores e orientam o armazenamento para o pesquisador (P1) e a coleta e uso dos dados publicados para os pesquisadores (P2, P3 e P4);
- têm informações convergentes e divergentes entre si, variando sua extensão e complexidade;
- fornecem requisitos para a elaboração dos PGDs pelos pesquisadores.

Os aspectos abordados no PGD dos repositórios variam de acordo com a comunidade alvo que atendem e aos tipos de dados que são armazenados, orientam os atores envolvidos na dinâmica dos repositórios e é primordial estarem em consonância com as orientações das agências financiadoras para a gestão de dados pelos pesquisadores. Observa-se que os PGDs não estão no mesmo nível de desenvolvimento e as informações variam entre si.

A preparação de um PGD envolve atividades em diferentes graus de formalidade, extensão, periodicidade, metas e objetivos. O desenvolvimento de seu conteúdo envolve as particularidades de cada área abrangida pelo repositório. As instruções inclusas no PGD variam dependendo dos objetivos e das características dos repositórios e dos conjuntos de dados neles depositados. Essas instruções devem ser claras para não gerar dúvidas e inseguranças.

As instruções sobre o PGD podem ser diversificadas dentro do mesmo repositório que cobrem áreas do conhecimento diferentes como no repositório da *University of California Los Angeles* UCLA que tem PGD para a *Sciences Guide*, *Humanities Guide* e outro para a *Social Science Guide*. Os PGDs dos repositórios traz benefícios pois fornecem uma compreensão das práticas na gestão dos dados e recomendações para os pesquisadores.

4 CONSIDERAÇÕES FINAIS

As universidades que instituíram repositórios de dados estão em grande número na Europa e Estados Unidos e as orientações e práticas de publicação e disponibilização de dados entre as comunidades estão presentes. No Brasil, com a orientação da Fapesp sobre a elaboração de PGD e a importância de compartilhar dados científicos, as universidades terão de implementar repositórios para o gerenciamento e publicação dos dados.

Os repositórios de dados fornecem infraestrutura que propicia a publicação e a divulgação de dados, a ampliação do impacto e da visibilidade das pesquisas, dos pesquisadores e das instituições das quais estão vinculados idealizando oportunidades adicionais de colaboração e conexão dentro e além de seu domínio.

Nesse contexto, implementar uma infraestrutura para apoiar os pesquisadores no ciclo de vida dos dados científicos, torna-se essencial. Os repositórios de dados científicos auxiliam no gerenciamento de dados e é fundamental que o gerenciamento seja planejado e estruturado desde o início de sua implementação. As diretrizes estabelecidas pelos repositórios estão descritas nos PGDs e podem levar em consideração as instruções das agências de fomento.

Plano de Gerenciamento de Dados é um documento que visa a explicitação da gestão de conjunto de dados e contribui para a gestão dos dados nos repositórios com instruções aos profissionais que neles trabalham orientando no gerenciamento por meio de diretrizes para a coleta, o armazenamento, a recuperação e o descarte, contribuindo ainda, no atendimento de requisitos relacionados a privacidade, a qualidade, a integração, a disseminação, aos direitos autorais e a preservação dos conjuntos de dados.

A gestão de dados é imprescindível para o bom andamento da pesquisa, porém, dentre as 55 universidades com Repositório de Dados, em 19 delas não foram localizados PGDs. As universidades implementaram o repositório de dados, no entanto, a gestão de dados ainda não está explicitamente evidenciada.

Os resultados da análise demonstram que das 100 melhores universidades do mundo, apenas 36 delas disponibilizam PDGs. Quando considerado que foram analisadas as

melhores universidades do mundo, esse fator comprova que ainda se tem um longo caminho para a conscientização da importância do PGD para a gestão de dados científicos.

Os benefícios gerados pela adoção dos PGDs são variados. Suas orientações ajudam a conduzir o gerenciamento dos dados pelos pesquisadores em suas pesquisas e nos repositórios assegurando que os diversos aspectos indicados pelas agências financiadoras sejam abordados e que os dados publicados tenham maior probabilidade de serem replicados.

Com a obrigatoriedade de elaboração de PGDs pelos pesquisadores e o encorajamento para disponibilização dos dados científicos publicamente pelas agências de fomento, os PGDs dos repositórios oferecem orientações e informações que auxiliam os pesquisadores na gestão de seus dados

As vertentes apresentadas corroboram que a área da Ciência da Informação, por meio do seu arcabouço teórico e prático, pode contribuir com a implementação de repositórios de dados, com o Ciclo de Vida dos Dados, com o gerenciamento dos dados e na orientação dos pesquisadores para elaboração de seus PGDs. As várias contribuições vão ampliando a atenção dada à gestão de dados por meio do estudo e fomento da utilização de PGDs. Todos os aspectos envolvidos na discussão estão emergindo no Brasil. Os profissionais da Ciência da Informação tem um campo estimado e produtivo de atuação e de pesquisa que irão contribuir para que os pesquisadores não comecem do zero na elaboração de seus PGDs, no gerenciamento dos dados e que as universidades terão profissionais engajados na implementação dos repositórios.

Como trabalhos futuros, abre-se a possibilidade de ampliar essa pesquisa para outras universidades e, também, instituições de pesquisa buscando estabelecer um mapeamento do processo de construção dos PGDs e de sua relação com as propostas apresentadas pelas agências de fomento. Esses estudos poderão inclusive subsidiar parâmetros globais que possam ser adotados de tal forma que a própria troca de informações entre os repositórios possa ser ampliada e sustentada por padrões de regras e normas que propiciem a configuração de diretrizes alinhadas.

REFERÊNCIAS

ALMEIDA, M. C. B. **Planejamento de bibliotecas e serviços de informação**. Brasília, DF: Brinquet de Lemos, 2005.

COX, A. M. PINFIELD, S. Research data management and libraries: current activities and future priorities. **Journal of Librarianship and Information Science**, London, v. 46, n. 4, p. 299-316, 2014. Disponível em: <<http://lis.sagepub.com/content/46/4/299.full.pdf+html>>. Acesso em: 27 set. 2016.

COUTO CORRÊA, F. **Gestión de datos de investigación**. Barcelona: Editorial UOC, 2016. Disponível em: <<http://bit.ly/2uwefAX>>. Acesso em: 2 jul. 2017.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais: **Chamada de Propostas de Pesquisa 2017. São Paulo, 2017a. Disponível em: <<http://www.fapesp.br/11068>>. Acesso em: 15 dez. 2017.**

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). Planos de gestão de dados se incorporam a projetos de pesquisa no Brasil. **São Paulo: Pesquisa FAPESP, 2017b. Disponível em: <<http://revistapesquisa.fapesp.br/2017/10/25/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/>>. Acesso em: 15 dez. 2017.**

FITZGERARD, A.; FITZGERARD, B.; PAPPALARDO, K. O futuro da política de dados. In: HEY, A. J. G. et al. **O quarto paradigma: descobertas científicas na era da eScience**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 209-216.

LEE, J. W. et al. DataNet: an emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. **AICHe Journal**, New York, v. 55, n. 11, p. 2757-2764, Nov. 2009. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/aic.12085/epdf>>. Acesso em: 05 jan. 2017.

KING, T.; FORCE, M. **Data Citation Index**. C2017. Disponível em: <<http://slideplayer.com/slide/5272632/>>. Acesso em 05 jan. 2017.

LEE, J. W. et al. DataNet: an emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. **AICHe Journal**, New York, v. 55, n. 11, p. 2757-2764, Nov. 2009. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/aic.12085/epdf>>. Acesso em: 05 jan. 2017.

MONTEIRO, E. C. S. A. **Direitos autorais nos repositórios de dados científicos: análise sobre os planos de gerenciamento dos dados**. 2017. 115 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017. Disponível em: <<http://hdl.handle.net/11449/149748>>. Acesso em: 30 abr. 2017.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Plano de Gerenciamento de Dados no contexto dos Repositórios de Dados de Universidades. In: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 1., 2017, Santa Catarina. **Anais eletrônicos...** Santa Catarina: UFSC, 2017. Disponível em: <<http://www.widat2017.ufsc.br/trabalhos-aceitos/>>. Acesso em: 30 nov. 2017.

OPENDOAR. **OpenDOAR charts**: Worldwide. 2017. Disponível em: <<http://www.opendoar.org/find.php?format=charts>>. Acesso em: 20 nov. 2017.

PINFIEL, S. Journals and repositories: an evolving relationship? **Learned Publishing**, v. 22, n. 3, p. 165-175, Jul. 2009. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1087/2009302/epdf>>. Acesso em: 05 jan. 2017.

PINFIELD, S; COX, A. M.; SMITH, J. Research data management and libraries: relationships, activities, drivers and Influences. **PLoS ONE**, v. 9, n. 12, p. 1-28, 2014. Disponível em: <<http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=fbf9aa61-6c97-445e-a489-9d76b4383641%40sessionmgr107&hid=128>>. Acesso em: 27 set. 2016.

PIWOWAR, H. A.; VISION, T. J. (2013) Data reuse and the open data citation advantage. **PeerJ**, San Diego, n. 1: e175 Disponível em: <<https://peerj.com/articles/175/#>>. Acesso em: 25 nov. 2017.

PLANOS de gestão de dados se incorporam a projetos de pesquisa no Brasil. **Revista Pesquisa FAPESP**, 2017. Disponível em: <<http://revistapesquisa.fapesp.br/2017/10/25/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/>>. Acesso em: 5 nov. 2017

PRYOR, G. (Ed.) **Managing research data**. United Kingdom: Facet Publishing, 2012. Disponível em: <<http://www.dcc.ac.uk/news/book-managing-research-data>>. Acesso em: 27 set. 2016.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação e informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>>. Acesso em: 20 out. 2016.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados científicos para bibliotecários de pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <http://carpedien.ien.gov.br:8080/bitstream/ien/1624/1/GUIA_DE_DADOS_DE_PESQUISA.pdf>. Acesso em: 5 out. 2016.

SESARTIC, A.; TOWE, M. Research data services at ETH-Bibliothek. **Journal IFLA**, Munich, v. 42, n. 4, p. 284-291, Nov. 2016. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0340035216674971>>. Acesso em: 5 out. 2017.

VAN LOAAN, J. E. et al. Quality evaluation of data management plans at a research university. **IFLA Journal**, Munich, v. 43, n. 1, p. 98-104, Mar. 2017. Disponível em: <<http://journals.sagepub.com.ez87.periodicos.capes.gov.br/doi/full/10.1177/0340035216682041>>. Acesso em: 2 dez. 2017.

WILLIAM, M.; BAGWELL, J.; ZOZUS, M. N. Data management plans: the missing perspective. **Journal of Biomedical Informatics**, San Diego, v. 71, p. 130-142, Jul. 2017. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1532046417300990?via%3Dihub>>. Acesso em 2 dez. 2017.

Editores do artigo: Enrique Muriel-Torrado, Edgar Bisset Alvarez, Camila Barros.