

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-8 – Informação e Tecnologia

DADOS CIENTÍFICOS: AS PRÁTICAS DE GESTÃO DOS PESQUISADORES BRASILEIROS NA CIÊNCIA DA INFORMAÇÃO

Renata Lemos dos Anjos - Universidade Federal da Paraíba (UFPB)

Guilherme Ataíde Dias - Universidade Federal da Paraíba (UFPB)

Adriana Alves Rodrigues - Universidade Federal da Paraíba (UFPB)

SCIENTIFIC DATA: THE MANAGEMENT PRACTICES OF BRAZILIAN RESEARCHERS IN THE INFORMATION SCIENCE AREA

Modalidade da Apresentação: Pôster

Resumo: A partir da hipótese de trabalho de que os dados provenientes de investigações científicas podem ficar armazenados nos notebooks, dispositivos externos dos pesquisadores, ou são descartados depois da publicação dos resultados em artigos científicos, tornando mais difícil a produção de novos estudos, esta pesquisa procura desvendar as práticas adotadas pelos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros no que diz respeito a gestão dos seus dados científicos. A presente pesquisa tem como objetivo geral identificar as dinâmicas do processo de gestão dos dados científicos adotadas por pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros. Encontra-se em andamento, e trata-se de uma pesquisa exploratória, de abordagem qualitativa e quantitativa para análise estatística dos dados que até então foram coletados através da aplicação do questionário, elaborado para fins desta pesquisa. Até o presente momento, dentre os 308 pesquisadores e sujeitos dessa pesquisa, 165 completaram o instrumento de pesquisa, e podemos perceber uma taxa de retorno de aproximadamente 53,57%. Conclui-se, parcialmente, que o presente estudo se mostra relevante diante do *e-Science* e no contexto da Ciência da Informação Brasileira, visto que traz um recorte da realidade da gestão dos dados científicos no cenário acadêmico brasileiro.

Palavras-Chave: Programas de Pós-Graduação em Ciência da Informação; Gestão de dados científicos; *e-Science*.

Abstract: Considering the working hypothesis that data from scientific investigations can be stored in the researchers' notebooks external devices, or discarded after publication of the results in scientific articles, making it difficult to produce new studies, this research seeks to unveil the practices of scientific data management applied by the researchers linked to the Brazilian Post-Graduate Programs in Information Science. The research main goal is to identify the dynamics related to the process of scientific data management adopted by researchers associated to the Brazilian Post-Graduate Programs in Information Science. This is an ongoing exploratory research, with a qualitative and quantitative approach to statistical analysis of the data collected through the deployment of a questionnaire, prepared for the purposes of this research. To date, within the 308 researchers selected for the research, 159 have completed the research instrument, currently the return rate is of approximately 51.62%. We can partially concluded that the present study is relevant to *e-Science*

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

and to the Brazilian Information Science context, since it brings a picture of the scientific data management in the Brazilian academic scenario.

Keywords: Graduate Programs in Information Science; Scientific data management; e-Science.

1 INTRODUÇÃO

Desde o momento em que foram publicados os primeiros periódicos científicos, a comunidade científica pauta-se por fazer ciência de maneira razoavelmente homogênea. Costumeiramente, as investigações são encerradas na publicação dos resultados alcançados em artigos de periódicos científicos (ANJOS, 2016). Esses resultados são gerados a partir da análise dos dados oriundos dos processos de investigação, os quais podem ser utilizados e reutilizados por outros pesquisadores na investigação de outros fenômenos, ou até para replicar um determinado experimento com fins de atestar a sua validade. Infelizmente, os dados oriundos de investigações científicas, chamados dados de pesquisa, ou, dados científicos, costumam ser descartados ou perdidos após a conclusão da pesquisa e da sua consequente publicação. Este cenário está mudando, visando a preservação destes dados e a origem de novas investigações, ao invés de um simples descarte, estes dados estão sendo coletados, gerenciados, preservados e compartilhados (BORGMAN, 2015).

Desta forma, a comunidade científica volta seu olhar para o processo de curadoria de dados que envolve a gestão atuante e a preservação dos dados de pesquisa durante todo o seu ciclo de vida. Em linhas gerais, o processo de curadoria de dados “envolve a gestão de dados de pesquisa desde o seu planejamento, assegurando a sua preservação por longo prazo, descoberta, interpretação e reuso” (SAYÃO; SALES, 2012, p. 179).

A iniciativa de realizar a preservação e reutilização dos dados científicos não está vinculada somente a ação de pesquisadores isolados, mas é um empreendimento que vem sendo desenvolvido de forma isolada ou em conjunto por instituições de pesquisas científicas. No Brasil, por exemplo, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) mantém um repositório digital de dados de projetos de pesquisa, o qual foi implementado através do *software* Dataverse, que objetiva a realização do compartilhamento, da preservação, por meio do arquivamento, desses dados para o mundo inteiro. O referido instituto além de promover a infraestrutura de *software* mencionada também está envolvido no desenvolvimento de uma política voltado para as questões que permeiam a emergência dos dados científicos.

Este novo cenário associado com a comunicação científica, em que, a “ciência é produzida a partir do uso, processamento, análise e compartilhamento de dados de

pesquisa” (GRAY, 2005 apud SALES *et al*, 2013), é também conhecido como o quarto paradigma científico, e denominado *e-Science*.

Nesta perspectiva, reflexões afloram das tendências contemporâneas da realidade da *e-Science*, com diferentes concepções de interação, compartilhamento de dados de pesquisa e colaboração entre os pares, nos fez indagar: **Quais as práticas adotadas pelos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros na gestão dos dados científicos?**

2 OBJETIVOS

No que tange ao objetivo geral, esta pesquisa busca **identificar as dinâmicas do processo de gestão dos dados científicos adotadas por pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros**. Para atender ao objetivo geral, foram delineados os seguintes objetivos específicos:

- a) Caracterizar os formatos de dados científicos utilizados pelos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros;
- b) Averiguar os tipos de dados científicos coletados pelos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros;
- c) Identificar as práticas de compartilhamento de dados empregadas pelos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros;
- d) Identificar as ações tomadas pelos pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros no sentido de armazenar e preservar os dados oriundos de pesquisas científicas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 E-SCIENCE

O fenômeno da *e-Science*, é caracterizado por uma ciência orientada a dados. Que cria oportunidades para o desenvolvimento da comunicação científica, uma vez que os dados de pesquisa são considerados fontes primárias de informação, quando compartilhados e acessíveis motivam o surgimento de novas investigações científicas.

Na *e-Science* nos deparamos com o apoio das Tecnologias Digitais de Informação e

Comunicação (TDIC) para com às atividades de investigação científica, como a colaboração no compartilhamento de dados e divulgação dos resultados (RIBES; LEE, 2010).

Em concordância com Sayão e Sales (2014, p. 79):

Esse fenômeno contemporâneo cria oportunidades sem precedentes para acelerar a pesquisa científica e gerar riquezas com base na exploração desse acúmulo de dados; abre a possibilidade de que a imensidão de dados gerados pela pesquisa científica contemporânea possa ser coletada, comparada e analisada, engendrando novos conhecimentos e novas questões de pesquisas. Ferramentas avançadas de software e de mineração de dados ajudam a interpretar e transformar os dados brutos em configurações ilimitadas de informação e conhecimento.

Conclui-se que a *e-Science* consiste no compartilhamento dos dados oriundos de investigações científicas, e em sua posterior análise, fazendo com que se tornem subsídios para novas informações, conhecimentos e questões de pesquisa (SAYÃO; SALES, 2014). Em linhas gerais, Oliveira e Silva (2016, p. 10), afirmam que “a *e-Science* se propõe a ser aberta, gratuita e de livre acesso”.

2.2 GESTÃO DE DADOS

Consoante ao que foi exposto, mediante a emergência da *e-Science*, pesquisadores estão dando especial atenção ao processo de gestão dos dados científicos, visando uma maior preservação, compartilhamento, e reuso dos mesmos.

O processo de gestão de dados, segundo Harvey (2010 apud COSTA; CUNHA, 2014, p. 202) consiste na:

(...) gestão ativa de dados durante o tempo em que ele continua a ser: acadêmico, científico, pesquisável, administrável e/ou de interesse pessoal, com objetivos de apoiar sua reprodutibilidade, reutilização e agregando valor a esses dados. Gerenciando-os do momento de sua criação, até que eles sejam determinados como não úteis e garantindo a sua acessibilidade em longo prazo, assim como a sua preservação, autenticidade e integridade.

Este processo objetiva a preservação dos dados de pesquisas, mantendo esses dados disponíveis por tempo indeterminado, inclusive, sobrepondo-se à obsolescência digital (software e hardware) (SALES; CAVALCANTI, 2015).

A prática de gestão de dados compreende o gerenciamento, a curadoria, e a preservação dos dados científicos durante todo o seu ciclo de vida, de forma que esses dados se tornem úteis para gerações de pesquisadores atuais e futuras.

4 PROCEDIMENTOS METODOLÓGICOS

Este tudo se caracteriza como uma pesquisa exploratória com análise qualitativa e quantitativa.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Diante da eclosão do *e-Science*, este estudo caracterizou-se como exploratório, visando uma melhor compreensão acerca do fenômeno mencionado. Este tipo de pesquisa é indicado quando se tem por finalidade ampliar o conhecimento sobre o objeto investigado, de forma a esclarecer conceitos e fornecer subsídios para as etapas subsequentes da investigação (VERGARA, 2006).

O presente estudo assume caráter quantitativo e qualitativo no que concerne às informações coletadas. Assume um caráter quantitativo, para realizar análise estatística dos dados obtidos através da aplicação do instrumento de pesquisa, questionário, e qualitativo para interpretação e análise dos dados coletados.

Para realização da pesquisa, foram selecionados os pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros, os quais foram convidados a colaborar com o presente estudo, mediante participação no questionário intitulado “A GESTÃO DE DADOS CIENTÍFICOS: Desvendando as práticas dos pesquisadores brasileiros na área da Ciência da Informação”, o qual foi elaborado e executado através do *software LimeSurvey*¹.

O instrumento de pesquisa foi composto por 23 questões, que foram distribuídas em 03 seções, da seguinte forma: seção introdutória, seção de identificação e, seção sobre a gestão de dados. Dentre as quais, 17 questões fechadas, 05 abertas e 1 mista. É válido destacar que algumas questões são correlacionadas, sendo dependentes de respostas anteriores, concluindo que alguns pesquisadores respondem todas as questões, e outros não.

No dia 01 de junho de 2017, foram criados os códigos de acesso com o objetivo de monitorar o acesso ao questionário, garantindo que os pesquisadores respondessem o questionário apenas uma vez. Desta forma, enviamos 308 convites para os pesquisadores vinculados aos Programas de Pós-Graduação em Ciência da Informação brasileiros, público-alvo deste estudo. A pesquisa encontra-se em andamento, igualmente ao questionário que permanece ativo para preenchimento.

Até o momento, obtivemos que dentre os 308 pesquisadores convidados, 182 aceitaram colaborar com este estudo, enquanto que 05, decidiram por não participar do estudo. Dos 182 que até o momento, aceitaram participar do estudo, 165 completaram o instrumento de pesquisa, e em contrapartida 17 pesquisadores não concluíram o questionário.

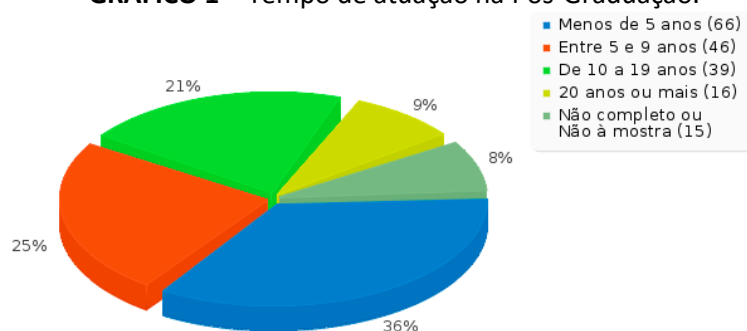
¹ <https://www.limesurvey.org/>

5 APRESENTAÇÃO E ANÁLISE DOS DADOS

Uma vez que o presente estudo se encontra em andamento, apresentaremos breves resultados provenientes da aplicação do questionário.

No que concerne ao perfil dos pesquisadores, componente da seção de identificação, até o momento temos, em sua maioria, com 36% os pesquisadores dos Programas de Pós-Graduação em Ciência da Informação brasileiros, possuem pouco vínculo de atuação no programa de pós-graduação, em média, menos de 5 anos. Posteriormente, 25% dos pesquisadores possuem entre 5 e 9 anos de atuação na pós-graduação. De 10 a 19 anos de atuação na pós-graduação temos 39 pesquisadores, um total de 21%. Com 20 anos ou mais, bem como, àqueles que não completaram o questionário temos 9%, como pode ser visto com valores aproximados no Gráfico 1.

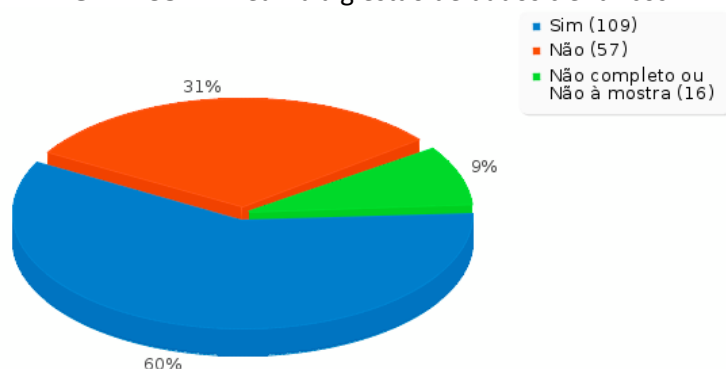
GRÁFICO 1 – Tempo de atuação na Pós-Graduação.



Fonte: Gráfico dos resultados elaborado pelo *LimeSurvey*.

No concerne a gestão dos dados, obtivemos, até o momento, que em sua maioria 109 pesquisadores afirmaram realizar a gestão dos dados provenientes de suas pesquisas, representando 60%. De maneira oposta, 57 pesquisadores não realizam a gestão dos dados científicos e encerraram nesta questão sua participação no questionário, visto que as questões seguintes diziam respeito as práticas referente a gestão de dados. Como apresentado na Gráfico 2, a seguir.

GRÁFICO 2 – Realiza a gestão de dados científicos.



Fonte: Gráfico dos resultados elaborado pelo *LimeSurvey*.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Na questão seguinte, questionamos aos pesquisadores que responderam de forma negativa a questão anterior, o motivo de não realizarem a gestão dos dados científicos. O desconhecimento acerca de tal prática foi a razão mais elencada pelos mesmos, dentre outras. Conforme expostas no Quadro 1.

QUADRO 1 – Razões para não fazer a gestão de dados.

"Considero ser muito recente e ainda não tenho muito conhecimento como fazê-lo."
"Falta de prática. Já li a respeito recentemente e é coisa nova."
"Desconheço métodos adequados de tratamento e armazenamento."
"Porque ainda tenho pouca familiaridade sobre esse procedimento. Preciso conhecer melhor as estratégias necessárias para o preparo e disponibilização dos dados de pesquisa nos repositórios."
"Por que nunca fui obrigada a fazer, e de fato, não sei onde começar! A única coisa que faço é guardar todos meus dados brutos (questionários, transcrições de entrevistas) em pastas - pastas físicas e digitais - estas últimas no meu micro e no Dropbox."
"Concluo a pesquisa e descarto os demais dados."
Falta de recursos para manter sistemas funcionando que permitam o acesso a dados. Tempo é um dos principais deles...
"Publico os resultados das pesquisas."
"Infelizmente ainda não faz parte das atividades relacionadas ao fazer científico."
Ainda não é uma tradição na nossa área. Guardei por vários anos, em fita magnética, os dados de minhas pesquisas. Entretanto, nunca recebi uma solicitação para compartilhar os dados! Assim, com o tempo, esses dados e o suporte de preservação ficaram defasados.
"Desinteresse."

Fonte: Dados da pesquisa, 2017.

Selecionamos estas duas questões de forma a proporcionar uma visão resumida dos resultados dos questionários, visto que a sua aplicação ainda não foi encerrada. O questionário, mesmo em andamento, já apresenta bons resultados de retorno, visto que dos 308 pesquisadores convidados a participar, obtivemos 165 questionários completos, o que representa aproximadamente 53,57% de retorno.

6 CONSIDERAÇÕES PARCIAIS

Ressalta-se que a pesquisa se encontra em andamento e que os dados ainda serão submetidos a uma análise estatística, sendo os resultados, apresentados de forma parcial.

Inferese que mesmo o presente estudo estando em andamento, o mesmo já mostra ser de grande relevância diante da realidade da *e-Science* e da Ciência da Informação Brasileira, tendo em vista que na questão referente à prática de gestão de dados, os pesquisadores, em sua maioria, afirmam realizar tal prática. Posto que, os mesmos são componentes dos quadros de discentes dos Programas de Pós-Graduação em Ciência da Informação de todo o Brasil, o estudo traz um recorte da realidade da gestão dos dados científicos na esfera acadêmica brasileira.

REFERÊNCIAS

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

ANJOS, R. L. **Desvendando as percepções e práticas adotadas pelos pesquisadores dos Programas de Pós-Graduação em Ecologia do Nordeste Brasileiro na gestão dos dados científicos.** Trabalho de Conclusão de Curso – Graduação em Biblioteconomia da Universidade Federal da Paraíba. 2016.

BORGMAN, C. L. **Big data, little data, no data: Scholarship in the Networked World.** London: MIT Press, 2015.

COSTA, M. M.; CUNHA, M. B. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação.** V. 19, n. 3. Belo Horizonte. Jul./Set. 2014. p. 189-206. Disponível em:
<<http://www.scielo.br/pdf/pci/v19n3/a10v19n3.pdf>>. Acesso em: 06 jul. 2017.

OLIVEIRA, A. C. S.; SILVA, E. M. Ciência aberta: dimensões para um novo fazer científico. **Informação & Informação**, v. 21, n. 2, p. 05-39, 2016. Disponível em:
<<http://www.brapci.ufpr.br/brapci/v/a/22211>>. Acesso em: 30 jul. 2017.

RIBES, D.; LEE, C. P. Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. **Computer Supported Cooperative Work**, v. 19, n. 3-4, p. 231-244, 2010. Disponível em: <
<http://static1.1.sqspcdn.com/static/f/1070098/14538475/1318114801180/Ribes+Lee+-+Cyberinfrastructure+Studies.pdf?token=0%2FF2mM1NTnwYvDutEQ0q1EAc1Qw%3D>>.
Acesso em: 30 jul 2017.

SALES, L. F.; CAVALCANTI, M. T. Seleção e avaliação de coleções de dados digitais de pesquisa: uma possível abordagem metodológica. **Informação & Tecnologia (ITEC): Marília/João Pessoa**, 2(2): 88-105, jul./dez., 2015. Disponível em: <
<http://periodicos.ufpb.br/index.php/itec/article/view/34134/17537>>. Acesso em: 30 jul. 2017.

SALES, L. F. et al. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-Science. In: XIV Encontro Nacional de Pesquisa em Ciência da Informação, 2013, Florianópolis. **Anais...** Florianópolis: XIV ENANCIB, 2013. Disponível em: <
<http://eprints.rclis.org/20666/1/APRESENTA%C3%87%C3%83O%20-%20ENANCIB%202013.pdf>>. Acesso em: 05 jul. 2017.

SAYÃO, L. F. SALES, L. F. Curadoria geral: um novo patamar para a preservação de dados digitais de pesquisa. **Informação & Sociedade.** V. 22, n. 3. João Pessoa. Set./Dez. 2012. p. 179-191. Disponível em:
<<http://www.ies.ufpb.br/ojs/index.php/ies/article/viewFile/12224/8586>>. Acesso em: 10 jul. 2017.

_____. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde.** V. 8, n. 2. Rio de Janeiro . 2014. p. 76-92. Disponível em: <<http://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>>. Acesso em: 05 jul. 2017.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração.** São Paulo: Atlas, 2006.