

CURADORIA E PRESERVAÇÃO DIGITAL: proposta de diretrizes para a gestão de dados científicos

CURATION AND DIGITAL PRESERVATION: guidelines proposed for research data management

Viviane Lílian dos Santos Barrozo 
Universidade Federal de Minas Gerais

Carlos Henrique Marcondes de Almeida 
Universidade Federal Fluminense

RESUMO

Esta pesquisa de mestrado, pretende elaborar um diagnóstico do estágio em que a UFMG encontra-se quanto à gestão de dados científicos, a conformidade às atividades e processos de curadoria digital e elucidar as diretrizes para a constituição de uma política de preservação digital. Uma pesquisa de abordagem qualitativa, natureza aplicada será explicativa e descritiva, se valerá da análise documental e bibliográfica. Como técnica de coleta de dados utilizará os questionários e entrevistas semiestruturados. Para obtenção dos dados da pesquisa, serão selecionados dois grupos de interesse, identificados como: "Produtores e Consumidores de Dados Científicos", composto por pesquisadores vinculados aos noventa cursos de pós-graduação nos níveis de mestrado e doutorado da UFMG e; "Formuladores de Políticas Institucionais", composto por pró-reitores, diretores das administrativas da UFMG. Espera-se promover a melhoria da infraestrutura informacional e de preservação digital, por meio da adoção de planos de gestão de dados e de preservação digital na UFMG.

Palavras-Chave: Dados Científicos, Política de Preservação, Princípios FAIR, Repositórios de dados.

ABSTRACT

This master's research aims to prepare a diagnosis of the internship in which UFMG is in relation to the management of scientific data, compliance to the activities and processes of digital Curation and elucidate the guidelines for the constitution of a digital preservation policy. A research of qualitative approach, applied nature will be explanatory and descriptive, will be used by documentary and bibliographic analysis. As a data collection technique, it will use questionnaires and semi-structured interviews. To obtain the research data, two interest groups will be selected, identified as: "Producers and Consumers of Scientific Data", composed of researchers linked to the ninety postgraduate courses at the master's and doctorate levels of UFMG and: "Institutional Policy Formulators", composed of pro-rectors, directors of UFMG administrative. It is expected to promote the improvement of information infrastructure and digital preservation, through the adoption of data management and digital preservation plans at UFMG.

Keywords: Scientific Data, Preservation Policy, FAIR principles, Data Repositories.

1. INTRODUÇÃO

Comunicar a ciência é algo secular, mas as transformações digitais ocorridas na segunda metade do Século XX, impactaram sobremaneira no fazer científico tanto no que tange a comunicação científica - disseminação dos resultados de pesquisa - quanto na divulgação científica - democratização do acesso à informação - e, suscitaram, a reflexão acerca dos processos informacionais para os ambientes digitais.

O intenso uso das tecnologias informáticas contribuiu para o aumento da produção de dados e, conseqüentemente, para o consumo de conteúdos digitais. Isto fez com que os dados atingissem um patamar de destaque para a ciência e reforçou a importância de gerir e preservar digitalmente, seja para acesso imediato ou a longo prazo. No contexto da Ciência Aberta, além da abertura dos periódicos, o fazer científico implica na abertura da pesquisa científica de forma ampla, incluindo, assim, a gestão de dados científicos ou curadoria de dados, as anotações e a revisão por pares, possibilitando o reuso de dados e a reprodutibilidade da pesquisa.

Nesse cenário, é possível observar que, a crescente relevância da gestão e preservação de dados digitais, o alinhamento entre processos e técnicas de curadoria e preservação de dados de investigação, bem como a participação institucional, são fundamentais no tratamento e na disseminação do conhecimento científico.

Neste trabalho, considera-se a gestão de dados científicos como um subsistema da área de Dados Abertos, uma das dimensões da Ciência Aberta que “visa tornar a pesquisa e os dados científicos acessíveis a todos” (UNESCO, online, 2021).

Dessa forma, o depósito dos conjuntos de dados resultantes das pesquisas produzidas no interior das universidades, denominada pequena ciência dá visibilidade e cumpre a função que, “resultados de pesquisa financiada por verbas públicas são um bem público e como tal devem ser disseminados em larga escala, resguardados os aspectos éticos, de privacidade, segurança e relativos à proteção de propriedade intelectual” (FAPESP, online, 2021).

Segundo Sayão e Sales (2019) na pequena ciência, por sua autonomia e potencial disruptivo, é onde ocorrem as inovações, as pesquisas interdisciplinares, as bases para a maioria dos artigos científicos e a conexão com os setores produtivos. Estes autores reafirmam que “na grande ciência, notadamente volta-se para os dados observacionais e referenciais, há um compromisso claro com a tríade curadoria, arquivamento e preservação, que inclui os

pressupostos de armazenamento seguro e ambientes confiáveis que garantam o acesso no tempo e no espaço”. Em contraste, na pequena ciência, o planejamento, a gestão, a preservação e a reutilização dos dados produzidos nessas pequenas dimensões, constituintes da cauda longa, apresentam-se desafiadores devido à sua heterogeneidade e singularidade. Portanto, para organizar o conhecimento científico deve-se observar as atividades e os processos que compõem todo o ciclo da investigação científica, sendo que a preservação dos dados é etapa antecessora à reutilização dos mesmos que é fundamental para o progresso da ciência.

Nesse sentido, as grandes ameaças ambientais ou crises pandêmicas como a recente resposta da comunidade científica à pandemia COVID-19 demonstrou muito bem como a Ciência Aberta pode acelerar a obtenção de soluções científicas para um desafio global (UNESCO, online, 2021).

Assim, a preservação dos dados digitais para acesso imediato ou a longo prazo deve constar do plano estratégico da instituição por meio da implementação de políticas, adoção de procedimentos e práticas que possibilitam gerenciar, continuamente, e salvaguardar seus acervos digitais.

1.2 Definição do problema

A gestão de dados científicos torna-se um desafio na perspectiva da Curadoria e Preservação Digitais. Pois, os conjuntos de dados resultantes das pesquisas brasileiras quando depositados em repositórios estrangeiros, bem como o armazenamento restrito aos laboratórios acadêmicos e nos equipamentos pessoais, dificultam a gestão integrada desses dados de pesquisa.

A ausência de políticas institucionais para preservar dados digitais a longo prazo, o risco de perda dos dados científicos devido à obsolescência tecnológica, fragilidade das mídias digitais e a falta de governança dos dados científicos produzidos no interior dos laboratórios acadêmicos, fazem com que, pesquisadores recorram aos repositórios de dados externos às suas instituições de filiação para depositarem os conjuntos de dados resultantes de pesquisa científica.

Segundo Sales e Sayão (2019) afirmam que grande parte desse estoque informacional não está disponível para o acesso, compartilhamento e reuso, pois, via de regra, permanecem

armazenados nos computadores pessoais dos seus criadores, sem nenhuma gestão que os tornem visíveis.

Em estudo recente, Dias; Dos Anjos e Araújo (2019) mencionam que há um uso corrente de repositórios de dados pela comunidade científica internacional. No mesmo estudo os autores apontam que pesquisadores dos programas de pós-graduação têm interesse em gerir os dados de suas pesquisas desde que haja uma formalização por parte de suas instituições.

Nesse sentido, o depósito dos conjuntos de dados brasileiros em repositórios estrangeiros, como também o armazenamento restrito aos laboratórios acadêmicos e, nos equipamentos pessoais, dificultam a gestão integrada dos dados de pesquisa.

Para Sayão e Sales (2012), parte considerável dos resultados das atividades de pesquisa está sendo criada em formatos digitais. E, embora os dados de pesquisa sejam de grande valor, estão sob risco de serem perdidos devido à obsolescência tecnológica e fragilidade das mídias digitais. Os autores salientam a importância da curadoria digital para 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 FAPESP (2021) evidencia que a gestão adequada dos dados de pesquisa é parte essencial das boas práticas de pesquisa. E, que dados resultantes de financiamento público devem ser gerenciados e compartilhados de forma a garantir o maior benefício possível para o avanço científico e tecnológico.

Além disso, a temática da pesquisa tem sido amplamente discutida em congressos, seminários e webinários nacionais e internacionais da área de Ciência da Informação durante os anos de 2020 e 2021. Essas discussões provocaram inquietações e questionamentos acerca da curadoria e preservação de dados.

Sendo a UFMG considerada uma das melhores universidades do Brasil, que desenvolve pesquisas em diversas áreas do conhecimento e, conseqüentemente, gera dados destas pesquisas. Este estudo visa endereçar a seguinte questão:

1. Quais políticas e planos de preservação digital para dados científicos são adotados na Universidade Federal de Minas Gerais (UFMG)?

1.3 Objetivos

Temos como objetivo geral, por meio de diretrizes, subsidiar a constituição de uma política de preservação de dados científicos na UFMG. Para isso, são elencados os objetivos específicos:

- diagnosticar o estágio em que a UFMG encontra-se quanto à gestão de dados científicos e, sua conformidade às atividades e processos de curadoria e preservação digital consagrados internacionalmente;
- elucidar as diretrizes para subsidiar uma política de preservação de dados científicos da/na UFMG.

2. FUNDAMENTAÇÃO TEÓRICA

Neste estudo, a gestão de dados científicos tem como ponto principal o gerenciamento e a preservação dos dados científicos produzidos no interior das Instituições de Ensino Superior (IES) brasileiras para que seja possível torná-los acessíveis, compartilháveis, reprodutíveis e reutilizáveis.

Para compreender, discutir e fazer propostas sobre a situação-problema apontada em 1.2, identificou-se como temas necessários ao desenvolvimento desta pesquisa: preservação digital; curadoria digital; dados de Pesquisa; Princípios FAIR e Repositórios de Dados.

Para contextualizar e elucidar acerca das políticas de preservação digital e planos de gestão de dados no Brasil e exterior, compor os quadros teóricos e conceituais quanto à gestão de dados de pesquisa, destacam-se as obras de Boeres (2004); Márdero Arellano (2008); Higgins (2011); Sierman et al. (2014); Sayão; Sales (2012;2013, 2016); Marcondes (2016); Siebra et al. (2016); Sales et al. (2017); Dias et al. (2019); Henning et al. (2019); Sales; Sayão (2019) e Sales et al. (2019).

Além do aporte teórico mencionado acima, o estudo será subsidiado por normas e recomendações de organizações como o Committee of Preservation and Conservation da IFLA (2020); Digital Curation Center - DCC (2020) e International Internet Preservation Consortium - IIP2 (2021). Assim como, por diretrizes e modelos propostos pelo Conselho Nacional de Arquivos - Conarq (2015); Open Access Infrastructure for research in Europe 4.0 - OpenAIRE Guidelines for Data Archives e Guidelines for Literature Repositories e PubFAIR - Confederation of Open Access Repositories - COAR (2019) com modelo

conceitual para uma estrutura modular e distribuída de publicação de código aberto; cOAlition S para o Acesso Aberto (2018) e a Matriz dos Níveis de Preservação Digital - LOP da NDSA (2019); Projetos como FORCE11 (2020); Facilitate Open Science Training for European Research FOSTER (2020); GO FAIR (2018) e InterPARES e dos estudos dos Grupos de Pesquisa DRÍADE/Rede Cariniana e Redes de Dados de Pesquisa Brasileira (RDP-Brasil).

3. METODOLOGIA

Com este estudo, pretende-se apresentar diretrizes para subsidiar a constituição de uma Política de Preservação de Dados Científicos para a UFMG. Para tanto, usará uma abordagem indutiva, generalizando suas conclusões a partir dos dados levantados.

Os procedimentos metodológicos se dividirão em 1-Teórico-metodológicos, efetivados por meio de revisão de literatura para viabilizar a compreensão das principais abordagens teóricas e conceitos referentes ao objeto de investigação e, se valerá do levantamento bibliográfico em bases de dados nacionais e internacionais com intuito de construir a revisão de literatura dos principais temas que compõem o objeto de pesquisa e, que propiciam inferir sobre o assunto tratado e; 2- Levantamento de dados através de pesquisa de campo, por meio de questionários semiestruturados.

A pesquisa é de natureza aplicada e abordagem qualitativa, pois tentará compreender as políticas de uma organização através das respostas dadas por seus formuladores. Quanto aos objetivos, será exploratória e explicativa, quanto aos procedimentos técnicos será bibliográfica, documental e pesquisa de campo, sendo que esta usará a técnica de coleta de dados por meio de survey - questionários e entrevistas semiestruturados - o primeiro, dirigido ao grupo de “Produtores e Consumidores de Dados Científicos”, este tendo como parâmetro a Matriz dos Níveis de Preservação Digital - LOP (2019), proposta pela NDSA, traduzida por um membro do grupo Dríade/Rede Cariniana. E, o segundo, dirigido ao grupo de “Formuladores de Políticas Institucionais”, abordando a governança de dados científicos para a implementação de processos e políticas na UFMG.

Os professores pesquisadores atuantes no noventa (90) cursos de pós-graduação nos níveis de mestrado e doutorado da UFMG serão identificados como “Produtores e Consumidores de Dados Científicos”. Os pró-reitores, diretores das administrativos, bem como os membros do Comitê Gestor do Repositório Institucional da UFMG, serão identificados como “Formuladores de Políticas Institucionais”. Esses pesquisadores serão contatados por meio de mensagens de correio eletrônico em suas contas institucionais. Os contatos dos

professores pesquisadores vinculados às unidades acadêmicas, os pró-reitores, dirigentes e tecnologistas serão obtidas nas informações disponibilizadas nas páginas dos órgãos aos quais os mesmos estavam vinculados. O instrumento para a coleta de dados será um questionário semiestruturado, devido ao fato da maior parte dos funcionários estarem em Home Office para atender ao protocolo de biossegurança da UFMG implantado a partir do cenário epidemiológico. O questionário online será disseminado por meio de uma conta institucional do Microsoft Forms®. A sua adoção justifica-se pela fácil tabulação das respostas. O projeto de pesquisa será submetido à apreciação e aprovação do Comitê de Ética em Pesquisa (COEP/UFMG). Preliminarmente, foram consultadas as fontes de informação a seguir: Base de Dados em Ciência da Informação (BRAPCI), Scientific Electronic Library Online (SciELO) e o Repositório Institucional Digital do IBICT (RIDI) para obtenção do aporte teórico. Além destas, serão consultadas outras bases internacionais, especializadas em Ciência da Informação e multidisciplinares como LISA, LISTA e *Web of Science*.

4. RESULTADOS PARCIAIS

Com os diagnósticos fundamentados pelas respostas aos questionários “Produtores e Consumidores de Dados científicos” e “Formuladores de Políticas Institucionais”, espera-se delinear o cenário da UFMG quanto à gestão de dados científicos, bem como as necessidades da observância às práticas de preservação digital. Além disso, com o resultado da revisão de literatura, pretende-se apresentar as diretrizes que poderão subsidiar a criação de políticas de preservação digital na UFMG.

5. CONSIDERAÇÕES FINAIS

Com as pesquisas bibliográficas e documentais foi possível perceber a emergência na melhoria da infraestrutura informacional e de preservação digital, adoção de planos de gestão de dados e planos de preservação digital, bem como a criação de repositórios de dados de pesquisa. Para isto, a literatura aponta a importância do envolvimento multidisciplinar para que as atividades de curadoria e preservação digitais sejam aplicadas aos acervos digitais. Além disso, os seminários e congressos que abordaram a temática, indicaram a utilização de consórcios para compartilhamento de soluções e redução dos custos para garantir o acesso à toda produção científica, incluindo os dados de pesquisa; atenção aos metadados de preservação e; conscientização acerca da importância de políticas

públicas de preservação digital. Espera-se contribuir para a gestão e preservação dos dados científicos produzidos nas IFES.

REFERÊNCIAS

DIAS, G. A.; DOS ANJOS, R. L.; ARAÚJO D. G. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reuso de dados. *Liinc em Revista*, v. 15, n. 2, 11 dez. 2019.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. Relatório de Atividades 2016. São Paulo: FAPESP, 2017. Disponível em <https://www.fapesp.br/openscience/>. Acesso em: 20 jan. 2021.

SALES, Luana Farias; CAVALCANTI, Márcia Teixeira. SELEÇÃO E AVALIAÇÃO DE COLEÇÕES DE DADOS DIGITAIS DE PESQUISA: uma possível abordagem metodológica. *Informação & Tecnologia*, v. 2, n. 2, p. 88-105, 2017.

SAYÃO, L. S. F.; SALES. (2016). Curadoria digital e dados de pesquisa. *AtoZ: novas práticas em informação e conhecimento*, 5(2), 67 – 71. Recuperado em: <http://dx.doi.org/10.5380/atoz.v5i2.49708>

SAYÃO, L. S. F.; SALES, L. F. Dados de pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 6, n. 1, 2013. Disponível em: <<http://www.brapci.inf.br/v/a/14157>>. Acesso em: 17 Jul 2018.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, v. 21, n. 2, p. 90-115, 2016.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital e dados de pesquisa. *AtoZ: novas práticas em informação e conhecimento*, [S.l.], v. 5, n. 2, p. 67-71, dez. 2016. ISSN 2237-826X. Disponível em: <<https://revistas.ufpr.br/atoz/article/view/49708/30179>>. Acesso em: 23 maio 2018. doi:<http://dx.doi.org/10.5380/atoz.v5i2.49708>.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para a preservação de dados digitais de pesquisa. *Informação & Sociedade*, v. 22, n. 3, 2012.

SOUZA, L. G. S.; AGANETTE, E. C. A preservação digital em longo prazo amparada por planos de ações: uma revisão sistemática de literatura. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, SP, v. 18, n. 00, p. e020027, 2020. DOI: 10.20396/rdbci.v18i0.8661185. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8661185>. Acesso em: 6 mai. 2021.

SOUZA, L. G. S.; AGANETTE, E. C. Preservação digital em Instituições de Ensino Superior: instrumentos para a oficialização e operacionalização da preservação de documentos arquivísticos digitais Tese (Doutorado em Ciência da Informação) – Departamento Organização e Tratamento da Informação da Universidade Federal de Minas Gerais, Brasília, 2020. Disponível em: <https://repositorio.ufmg.br/handle/1843/34550>. Acesso em: 6 maio. 2021.

UNESCO. Open Science. 2021. Disponível em: <https://en.unesco.org/science-sustainable-future/open-science> . Acesso em: 30 abr. 2021.

