

Curadoria digital de dados de pesquisa: um levantamento qualitativo acerca da produção científica

Sabrina Beatriz Martins Andrade¹, Caterina Marta Groposo Pavão²

RESUMO

Para tratar do volume expressivo, no que tange aos dados de pesquisa, a curadoria digital surge como uma ferramenta passível de tratar de dados a fim de que os mesmos possam estar disponíveis e interoperáveis, facilitando o seu acesso permanente além de contribuir para novas e futuras pesquisas por meio da melhoria da visibilidade desses dados. Este estudo apresenta uma abordagem qualitativa acerca da produção científica, dos últimos 10 anos, na América Latina, sobre curadoria digital de dados de pesquisa, com o objetivo de identificar o que tem sido produzido sobre o tema. As bases de dados selecionadas para o estudo permitiram obter um panorama da quantidade de produção científica e a necessidade de incremento de estudos sobre a temática, no âmbito latino-americano. Notou-se, também, o quanto é importante uma padronização nos campos de busca e filtros disponibilizados nas bases de dados, a fim de garantir uma recuperação satisfatória e consequentemente a sustentação de um referencial teórico.

Palavras-chave: Curadoria digital; Curadoria de dados; Dados de pesquisa.

INTRODUÇÃO

As tecnologias digitais e da *web* avançam rapidamente. A sociedade contemporânea produz e consome informações em formato digital em um volume expressivo. Por meio desses registros digitais criados e reproduzidos em todos os segmentos da sociedade, a mudança no comportamento, negócios e ensino são notáveis e acabam contribuindo para o surgimento de novos fenômenos, como o *Big Data*, o qual consiste em conjuntos de dados, caracterizados pela variedade, volume e velocidade de informações.

No âmbito científico as transformações ocorrem naturalmente acompanhando a evolução desenfreada das informações e dos dados digitais. Novas formas de representar e organizar a produção científica surgem junto com a necessidade de novas metodologias de produção de conhecimento científico. Deste modo, a ciência aberta, a qual considera o conhecimento científico um bem da humanidade, alia-se a esse cenário demandando que metodologias, dados, instrumentos e *softwares* estejam abertos, ou seja, totalmente acessíveis garantindo a reprodutibilidade e autocorreção da ciência e transparência de seus fluxos. (SAYÃO; SALES, 2016).

Frente a essas mudanças, os dados científicos passam a ser os protagonistas das pesquisas e necessitam de tratamento desde a sua criação, planejamento, preservação, além do uso e reuso, e sua reprodutibilidade. A curadoria digital garante esse papel, a fim de que os dados possam estar disponíveis e interoperáveis, facilitando o acesso permanente aos dados e contribuindo para novas e futuras pesquisas.

¹ Mestrando do Programa de Pós-Graduação em Ciência da Informação – PPGCIN/UFRGS
sabinabibliofurg@gmail.com

² Professora do Programa de Pós-Graduação em Ciência da Informação – PPGCIN/UFRGS;
caterina@cpd.ufrgs.br

Os dados de pesquisa precisam ser organizados e estruturados com a finalidade de permitir seu armazenamento em formatos adequados, preservação e acesso, conseqüentemente garantindo sua confiabilidade, autenticidade e integridade (GRÁCIO; ARELLANO, 2020).

Para tanto, a curadoria digital, envolve ações que precisam ser executadas regularmente, objetivando a garantia de que o documento digital seja preservado de forma segura, mantendo a autenticidade, integridade, confiabilidade, originalidade, organicidade, acessibilidade e preservação do documento digital a longo prazo (SILVA; SIEBRA, 2019).

O termo foi utilizado pela primeira vez em 2001 no evento *Digital Curation: Digital Archives, Libraries and e-Science Seminar*. Logo, em 2002, foi fundada a *Digital Preservation Coalition (DPC)*, uma empresa sem fins lucrativos criada como um consórcio entre várias agências do Reino Unido e Irlanda com interesse na preservação da informação digital. E foi em 2004 que inaugurou o *Digital Curation Centre (DCC)*, responsável pela elaboração de recomendações, normas e políticas de gestão para acervos científicos digitais no Reino Unido. Focado no desenvolvimento de capacidade e habilidades para o gerenciamento de dados, propôs o mais conhecido modelo de ciclo de vida da curadoria digital, o *Digital Curation Lifecycle Model*, que abrange uma sequência de atividades como: conceituar; criar e receber; avaliar e selecionar; capturar; ação de preservação; armazenar; acessar, usar e reusar; transformar; eliminar; reavaliar e; migrar.

No Brasil, conta-se com a Rede Cariniana, conhecida por meio de uma iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) surgida da necessidade de criar uma rede de serviços de preservação digital de documentos eletrônicos brasileiros, com o objetivo de garantir seu acesso contínuo a longo prazo.

Abbott (2008), uma das autoras mais citadas na literatura acerca do tema, menciona que a curadoria pode ser definida como a gestão ativa e preservação de objetos digitais a fim de promover acesso, uso e reuso a longo prazo. Ainda, aponta que a curadoria considera como parte de suas atividades os processos para gestão, manutenção, preservação e agregação de valor aos objetos digitais. Dessa forma, assegura a sustentabilidade dos dados para uso futuro e emprega tecnologias que potencializam a qualidade dos documentos digitais.

Diante o exposto, por se tratar de um assunto que está desenvolvendo-se gradualmente no país, neste estudo realizamos um levantamento quali-quantitativo da produção científica dos últimos 10 anos por meio de uma busca, em recursos informacionais, sobre o tema curadoria digital de dados de pesquisa, com o objetivo de identificar o que tem sido produzido acerca do assunto.

METODOLOGIA E PRINCIPAIS DESCOBERTAS

Para atender o objetivo proposto realizou-se uma pesquisa bibliográfica em ferramentas que abrangem o campo da Ciência da Informação e que julga-se melhor abranger o que a academia vem estudando, tais como: a Biblioteca Digital de Teses e Dissertações (BDTD), a Base de Dados Referenciais de Artigos de Periódicos em

Ciência da Informação (BRAPCI), o Google Acadêmico, a Rede de repositórios institucionais de acesso aberto (La Referencia), o Portal de publicações científicas em acesso aberto (Oasisbr), o Portal de Periódico da CAPES e o Repositórios Científicos de Acesso Aberto de Portugal (RCAAP).

Determinou-se o intervalo entre os anos de 2011 e 2021, filtrando a busca pelos campos de assunto ou palavra-chave e resumo, e em casos que não houve essa possibilidade, pelos campos de texto completo e/ou título. Foram utilizados os seguintes termos: “curadoria digital” e “dados de pesquisa” na língua portuguesa, “digital curation” e “research data curation” na língua inglesa, “curación digital” e “datos de investigación” na língua espanhola, adequando-os às diferentes ferramentas de busca e utilizando operadores booleanos sempre que possível e/ou necessário.

A tabela abaixo apresenta os resultados obtidos nas buscas realizadas nas bases de dados, onde é possível observar os termos utilizados, bem como os campos de busca utilizados de acordo com o que cada base disponibiliza.

Tabela 1. Resultados quantitativos da pesquisa

FERRAMENTA DE BUSCA	TERMO UTILIZADO	CAMPO DE BUSCA	Nº DE ITENS RECUPERADOS
BDTD	“curadoria digital” e “dados de pesquisa”	Assunto	03
	“curadoria digital” e “dados de pesquisa”	Resumo	05
BRAPCI	“curadoria digital” e “dados de pesquisa”	Palavra-chave	02
	“curadoria digital” e “dados de pesquisa”	Resumo	00
	“curadoria digital” e “dados de pesquisa”	Texto completo	16
GOOGLE ACADÊMICO	“curadoria digital” e “dados de pesquisa”	Em qualquer lugar do artigo	402
	“curadoria digital” e “dados de pesquisa”	No título do artigo	03
	“digital curation”	Em qualquer lugar do artigo	14.800
	“digital curation”	No título do artigo	638
	“research data curation”	Em qualquer lugar do artigo	9.960
	“research data curation”	No título do artigo	64
	“curación digital”	Em qualquer lugar do artigo	108
	“curación digital”	No título do artigo	06
	“datos de investigación”	Em qualquer lugar do artigo	15.300
“datos de investigación”	No título do artigo	210	
LA REFERENCIA	“curación digital”	Materia	01
	“datos de investigación”	Materia	34
OASISBR	“curadoria digital” e “dados de pesquisa”	Assunto	10
	“curadoria digital” e “dados de pesquisa”	Resumo	23
PORTAL DA CAPES	“digital curation”	Assunto	444
	“research data curation”	Assunto	03
RCAAP	“curadoria digital” e “dados de pesquisa”	Assunto	02

	“curadoria digital” e “dados de pesquisa”	Texto integral	43
	“digital curation”	Assunto	31
	“digital curation”	Texto integral	27
	“research data curation”	Assunto	04
	“research data curation”	Texto integral	50

Fonte: Dados da pesquisa, elaborado pelas autoras.

É interessante observar os resultados da pesquisa no Google Acadêmico, em que somente é possível realizar a busca no texto completo ou no título do artigo. Nota-se a discrepância entre os resultados das bases de dados, percebe-se que este recurso não apresenta uma estratégia de busca confortável, pelo fato de não fornecer uma variedade de filtros satisfatórios que garantam uniformidade de pesquisa ou comparações com outras bases de dados. A recuperação em todo o texto do documento, mesmo utilizando aspas para delimitar os termos compostos, sempre terá a tendência de recuperar mais registros, que muitas vezes não são representativos para um estudo mais específico.

Quanto à Rede de Repositórios La Referencia, que abrange publicações científicas da América Latina, em acesso aberto, é possível identificar que por meio do assunto “curación digital” recuperou-se somente um documento, já com o assunto “datos de investigación”, obteve-se 34 documentos. O que nos leva a refletir que a temática dados de pesquisa pode ser mais presente nas produções científicas, mas o tema curadoria digital ainda não despertou tanto interesse.

No Portal de Periódicos da CAPES percebe-se a diferença de documentos recuperados com o termo “digital curation” (444 itens) para o termo “research data curation” (três itens). Em um primeiro momento, pode-se pensar que o assunto “curadoria de dados de pesquisa” não está sendo tão abordado quanto o assunto “curadoria digital”. Por outro lado, o resultado pode estar relacionado ao fato da curadoria digital abranger qualquer tipo de dados e/ou materiais digitais e por isso ter uma quantidade maior de documentos recuperados. Também, pode ser a forma como os pesquisadores referem-se à curadoria de dados de pesquisa, fazendo uso de um termo mais amplo. Estas hipóteses somente poderão ser confirmadas na continuação do estudo quando cada um dos documentos recuperados serão analisados.

No RCAAP, a busca com os mesmos termos em inglês, usados também, no campo assunto, resultou em: “digital curation” (31 itens) e “research data curation” (quatro itens). Poderia-se fazer que uma comparação com os resultados obtidos no Portal da CAPES, não em relação a número absolutos recuperados em cada ferramenta, pois o Portal da CAPES oferece recuperação em diversas bases de dados de publicações periódicas internacionais, enquanto o Portal de Repositórios de Portugal contempla somente conteúdos científicos de repositórios institucionais das entidades nacionais de ensino superior. Mas, a comparação está pautada em relação à discrepante proporção de documentos recuperados com a mesma estratégia de busca, “digital curation”, 444 itens no Portal da CAPES e 31 no RCAAP e para o termo “research data curation”, três itens Portal a CAPES e quatro no RCAAP. Mais uma vez, esta relação só poderá ser entendida na continuação do estudo.

Neste primeiro momento não foram analisados os documentos recuperados, o total de itens relevantes, a necessidade de ampliar ou restringir os termos e os filtros utilizados para aprimorar a busca e nem foram excluídas a duplicidade de documentos recuperados, que de antemão sabe-se que ocorre visto que algumas das ferramentas consultadas são apenas provedoras de serviços, ou seja, coletam informações de provedores de conteúdos.

Em relação às bases de dados de modo geral, chama a atenção a diferença dos campos de busca. A falta de padronização, de campos de busca e filtros disponíveis, dificulta as comparações dos resultados recuperados, visto que podem-se perder itens relevantes, assim como recuperar itens não relevantes. Isso reforça os conceitos de exaustividade, especificidade, revocação e precisão na indexação. Em que, a exaustividade é a aplicabilidade em número suficiente de termos para abranger o conteúdo temático principal do documento, ou seja, quanto mais termos forem aplicados para indexar um documento, mais acessível ele se tornará e, conseqüentemente, mais vezes poderá ser recuperado” (LANCASTER, 2004).

Assim, percebe-se que quanto mais exaustiva for a busca, maior será o índice de revocação e em contrapartida menor será a precisão.

Segundo Nóbrega (2011), enquanto o índice de revocação é medido pela razão entre os documentos relevantes recuperados e o total de documentos relevantes existentes no sistema, o índice de precisão é calculado pela razão entre os documentos relevantes recuperados e o total de documentos.

Portanto, entendemos que a revocação e a precisão tratam-se de índices que avaliam a qualidade da resposta do sistema, sendo inversamente proporcionais. Logo, quanto maior a precisão de um sistema, menor será sua revocação.

Assim, é fundamental que haja uma eficiente arquitetura nos sistemas de buscas visando qualificar as solicitações de pesquisa dos usuários, além disso, faz-se necessário uma padronização tanto nacional quanto internacional nestes sistemas, a fim de que se possa ter um resultado homogêneo e satisfatório para as análises de pesquisas futuras.

CONSIDERAÇÕES FINAIS

A Curadoria Digital tem emergido na literatura como resultado dos impasses provenientes da informação digital. Essa informação, que vem desenvolvendo-se e sofrendo transformações em seu formato, volume, disponibilidade e acesso em meio ao ambiente interativo e dinâmico das tecnologias de informação e comunicação.

Quanto aos recursos informacionais utilizados na pesquisa, para a busca do que vem sendo publicado nos últimos 10 anos sobre curadoria digital de dados de pesquisa, nota-se o quanto é importante uma padronização nos campos de busca para que possa garantir uma recuperação satisfatória a fim de sustentar um referencial teórico ou estado da arte sobre um tema específico. As bases de dados selecionadas para o estudo são fundamentais para obter-se um panorama sobre a produção científica acerca do tema proposto. Em estudo futuro será analisada a frequência e a evolução da temática ao longo do tempo, entre outros aspectos.

Quanto à curadoria digital, conclui-se, que ela contribui para garantir validade aos dados de pesquisa. Além disso, ela proporciona a utilização de padrões comuns entre diferentes conjuntos de dados, criando assim mais oportunidades em buscas transversais de colaboração. Do ponto de vista financeiro, o compartilhamento, o reuso de dados e oportunidades de novas análises protegem e valorizam o investimento inicial para obtenção dos dados.

Por fim, a curadoria digital proporciona a sustentabilidade dos dados para o futuro. A tecnologia aliada aos recursos estratégicos e metodológicos utilizados na curadoria digital garantem e facilitam o acesso contínuo aos dados digitais por meio da melhoria de qualidade.

Ressalta-se, finalmente, que este estudo refere-se a uma primeira aproximação com a temática que será abordada durante a pesquisa, durante o mestrado. Foi possível identificar outras ferramentas nas quais as buscas, também, deverão ser realizadas, assim como termos que deverão ser incluídos para melhor representar os estudos de curadoria de dados de pesquisa.

REFERÊNCIAS

ABBOTT, Daisy. **What is digital curation?** Edinburgh, UK: Digital Curation Centre, 2008.

DCC. Digital Curation Centre. [site]. Disponível em: <https://www.dcc.ac.uk/>. Acesso em: out. 2021.

DPC. Digital Preservation Coalition. [site]. Disponível em: <https://www.dpconline.org/>. Acesso em: out. 2021.

GRÁCIO, J. C. A.; ARELLANO, M. A. M. A gestão da preservação digital de dados de pesquisa proposta de um modelo processual. **Revista Brasileira de Preservação Digital**, Campinas, v. 1, e020001, 2020. Disponível em: <https://brapci.inf.br/index.php/res/download/158987>. Acesso em: jul. 2021.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004.

NÓBREGA, D. L. **Indexação de artigos de periódicos em Ciência da Informação: elaboração de política de indexação para a base ABCDM**. 2011. 106f. Monografia (Bacharelado em Biblioteconomia) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2011.

REDE CARINIANA. Rede Brasileira de Serviços de Preservação Digital. IBICT. [site]. Disponível em: <https://cariniana.ibict.br/>. Acesso em: out. 2021.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: Novas Práticas em Informação e Conhecimento**, v. 5, n. 2, p. 67-71, 2016. Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: jul. 2021.

SILVA, F. M. O.; SIEBRA, S. A. Desmaterializando o documento: contribuições da diplomática para a curadoria digital de documentos arquivísticos digitais. **Archeion Online**, João Pessoa, v. 6, n.2, p. 23-44, jan./jun. 2019.