

Atuação bibliotecária em repositórios de dados de pesquisa: da perspectiva da entrega à perspectiva do acesso

The librarian's role in research data repositories: from the delivery perspective to access perspective

Roberta Cristina Dal'Evedove Tartarotti (1), Paula Regina Dal'Evedove (2), Mariângela Spotti Lopes Fujita (3)

(1)(3) Universidade Estadual Paulista (UNESP), Av. Higino Muzi Filho, 737, Marília, SP, roberta_tartarotti@yahoo.com.br. (2) Universidade Federal de São Carlos (UFSCar), Rodovia Washington Luís, km 235, São Carlos, SP, dalevedove@ufscar.br. (3)mariangelaslf57@gmail.com

Resumo

Os repositórios de dados de pesquisa emergem como sistemas contemporâneos e heterogêneos de recuperação da informação científica, em que se destacam dois conceitos: *Biblioteconomia de dados* e *bibliotecário de dados*. Essas novas concepções se atrelam às discussões em Organização do Conhecimento na defesa de instituições eficazes e sustentáveis, em atenção aos métodos e práticas empregadas para a análise de dados de pesquisa, especialmente no ambiente web, tornando o papel do bibliotecário na organização destes sistemas muito mais desafiadora. Diante desse cenário, objetiva-se realizar pesquisa exploratória sobre os aspectos relacionados a esses novos conceitos no contexto das bibliotecas universitárias e à perspectiva que se apresenta para os estudos de organização da informação. Observa-se que os repositórios de dados de pesquisa trazem uma perspectiva inovadora à atuação do bibliotecário ao demandarem a observância de questões relacionadas à cobertura do sistema, metadados, acesso e qualidade dos dados, preservação digital, confiabilidade dos sistemas e interoperabilidade, os quais exigem o envolvimento de pesquisadores e profissionais na condução de pesquisas e ações dedicadas à gestão digital de dados de pesquisa e às perspectivas e desafios que se apresentam para a organização da informação.

Palavras-chave: Dados de pesquisa; Repositório de dados; Atuação bibliotecária. Biblioteconomia de dados. Bibliotecário de dados; Organização da informação.

Abstract

The research data repositories emerge as contemporary and heterogeneous systems of information retrieval, in which two concepts stand out: *Data librarianship* and *data librarian*. These new conceptions refer to the discussions of the Knowledge Organization area in defense of effective and sustainable institutions, complying with the methods and practices used for the analysis of research data, especially in the web environment, making the librarian's role in the organization of these information retrieval systems much more challenging. In view of this scenario, the objective is to conduct exploratory research on the aspects related to these new concepts in the context of university libraries and the perspective that is presented for studies of information organization. It is observed that research data repositories bring and innovative perspective to the librarian's role, demanding the fulfillment of issues related to system coverage, metadata, data access and quality, digital preservation, system reliability and interoperability, which require the involvement of researchers and professionals. conducting research and actions dedicated to the digital management of research data and the perspectives and challenges presented for the organization of information.

Keywords: Research data; Data repository; Librarian's role; Data librarianship; Data librarian; Information organization.

1 Introdução

Ao longo da história da Biblioteconomia e da Ciência da Informação, diversas foram as técnicas, padrões e instrumentos desenvolvidos para organizar, recuperar e compartilhar informações. Esse esforço contínuo deve-se à complexidade e desafios que se apresentam à organização do conhecimento e da informação em uma sociedade tecnológica e interconectada em rede.

Em um ambiente tradicional de biblioteca universitária, por exemplo, os metadados se concentravam na descrição de documentos físicos (mais comumente livros e periódicos), com informações sobre o autor, título, editor, data de publicação, cabeçalhos de assunto e número de classificação. Considerando-se que o processamento de máquina ainda era desconhecido,

este conjunto de descrições foi originalmente projetado para uso humano (RUBIN, 2016, p. 339). Todavia, com o advento e proliferação da Internet e a presença do digital, decorrência do avanço tecnológico e de uma sociedade cada vez mais imersa no ciberespaço, surgem modernos e inovadores sistemas de recuperação de informação. Por decorrência, novas ferramentas, padrões e modelos de dados desenvolvidos para organizar e gerenciar recursos informacionais digitais são projetados, tornando sua organização muito mais desafiadora (CHOWDHURY; CHOWDHURY, 2007, p. 221).

Os dados de pesquisa surgem nesse cenário de revolução como recursos essenciais e estratégicos na produção científica, sendo requeridos métodos que viabilizem o compartilhamento e o reúso de resultados

provenientes de comunicações científicas. Dados de pesquisa são definidos como “aquilo que é coletado, observado ou criado em formato digital, por propósitos de análise para produzir resultados originais de pesquisa¹”, tais como questionários, protocolos experimentais, cadernos de laboratório, gráficos ou informações sobre as configurações de um determinado equipamento, dentre outros.

O êxito do movimento de compartilhamento dos dados de pesquisa está diretamente relacionado à (re)configuração dos métodos, técnicas e tecnologias empregadas na geração de produtos e/ou serviços em sistemas de informação científica.

As mudanças em arquivos de dados nacionais centralizados, principais locais de preservação e disseminação de dados de pesquisa bruta, levaram ao desenvolvimento dos chamados *repositórios de dados de pesquisa*, que se propõem a manter e compartilhar dados de pesquisas produzidos dentro de uma determinada universidade, seguindo o movimento de acesso aberto, em que as cópias de publicações estão sendo coletadas e arquivadas em nível institucional.

Os repositórios de dados de pesquisa abrangem os *conjuntos de dados de pesquisa*, “coletados como parte de um projeto de pesquisa, descritos e disponibilizados para uso por outros pesquisadores”, que não se limita à descoberta de informações e acesso, exigindo um envolvimento mais profundo com o processo de pesquisa (RICE; SOUTHALL, 2016, p. 16, tradução nossa).

Por parte dos editores científicos, há uma tendência “em exigirem que os dados subjacentes das publicações sejam preserváveis, detectáveis e acessíveis em um repositório, com links recíprocos entre o artigo publicado originalmente e os dados” (RICE; SOUTHALL, 2016, p. 15, tradução nossa). Logo, os repositórios digitais de dados de pesquisa surgem para apoiar a produção, armazenamento, uso, reúso, acesso e compartilhamento dos conjuntos de dados gerados durante as várias etapas do processo da pesquisa científica.

No ambiente web, outros atores contribuem com a descrição de metadados ao conteúdo digital, dentre os quais temos os produtores de bancos de dados, criadores de sites, editores, fornecedores, organizações científicas e técnicas, desenvolvedores de software e usuários gerais e especialistas. Entretanto, os bibliotecários continuam desempenhando um importante papel na criação de metadados, porém, como parte de um domínio muito maior de criadores de metadados, que não apenas criam, mas contribuem com padrões e práticas relacionados aos metadados (RUBIN, 2016, p. 339).

No contexto dos repositórios de dados de pesquisa como sistemas contemporâneos e heterogêneos de

recuperação da informação emergem dois conceitos: *Biblioteconomia de dados* e *bibliotecário de dados*. Tais concepções estão em consonância com as discussões no campo da Organização do Conhecimento referentes à defesa de instituições eficazes e sustentáveis, tendo em vista a necessidade do estabelecimento de métodos e práticas específicos para a análise de dados de pesquisa, especialmente no ambiente web.

Isso posto, o trabalho apresenta resultados de uma pesquisa exploratória na literatura nacional e internacional sobre os aspectos relacionados a esses novos conceitos no contexto das bibliotecas universitárias e à perspectiva que se apresenta para os estudos de organização da informação, tomando-se como base os ideais apresentados por Rice e Southall (2016). A justificativa que se coloca para a presente investigação decorre da incipiente discussão dessa conjuntura na Ciência da Informação brasileira, com atenção especial à atuação do bibliotecário de dados, o desenvolvimento de competência em dados e os métodos e práticas aplicados para viabilizar o uso e consumo de dados de pesquisa.

2 Biblioteconomia de dados: da perspectiva da entrega à perspectiva do acesso

O conjunto das atividades de uma biblioteca universitária abrange procedimentos que visam permitir que os recursos informacionais, tanto impressos quanto digitais de uma determinada coleção, sejam organizados e disponibilizados à comunidade acadêmica.

No ambiente acadêmico, os bibliotecários de dados tornam-se especialistas em informação científica, capazes de reunir e assessorar no desenvolvimento de coleções de conjuntos de dados de pesquisa, uso e preservação dos dados, como também no acesso a dados já arquivados ou na criação de novos dados, como parte de sua prática de pesquisa. Nesse sentido, “bibliotecas e seus bibliotecários de dados são idealmente colocados para atender a essas necessidades”, criando o que Rice e Southall (2016, p. 16) denominam de “um novo mapa de suporte e serviços para pesquisadores”. Surge, então, o termo *Biblioteconomia de dados*, perspectiva que coloca a área como protagonista no desenvolvimento de produtos e serviços de informação científica relacionados ao uso e consumo de dados de pesquisa no ambiente web. Nas palavras dos referidos autores,

A Biblioteconomia começa então a ser entendida não apenas como algo que apóia a descoberta e o acesso a títulos publicados ou recursos informacionais, mas também como algo que se envolve com a *conduta* de pesquisa e investigação acadêmica (RICE; SOUTHALL, 2016, p. 2, tradução nossa, grifo dos autores).

Por ser um conceito novo na área, poucos são os estudos dedicados ao tema. Na Ciência da Informação brasileira, é escassa a produção de pesquisas que considerem dados de pesquisa e repositórios digitais de dados de pesquisa como tema de investigação, podendo-se citar os trabalhos desenvolvidos por Sayão e Sales (2016), Bertin, Visoli e Drucker (2017), Rocha, Sales e Sayão (2017), Cavalcanti e Sales (2017), Santos e Rockembach (2018), Semeler e Pinto (2019), dentre outros.

Questões específicas dos repositórios digitais de dados de pesquisa foram amplamente trabalhadas por Sayão e Sales (2016). Para tanto, os autores apresentaram um panorama das principais características, categorias, benefícios, funções e infraestruturas necessárias aos repositórios de dados de pesquisa para darem sustentação ao acesso e à gestão de coleções de dados e dados de pesquisa. Na prática, a exigência recai para uma infraestrutura tecnológica e gerencial que possibilite que os dados de pesquisa sejam assistidos durante todo o seu ciclo de vida, de forma permanente e sustentável. Concomitante a essa questão, Cavalcanti e Sales (2017) debruçam-se sobre a gestão de dados de pesquisa e elegem a União Europeia como berço para o levantamento quantitativo de repositórios de dados de pesquisa existentes nestes países. A contribuição do estudo dá-se com a apresentação de um panorama das políticas da gestão de dados de pesquisa em repositórios europeus.

Ao considerarem os cadernos eletrônicos de laboratório como alternativa aos cadernos de laboratório em papel, Rocha, Sales e Sayão (2017) explanam sobre o importante papel do bibliotecário como mediador da preservação e do livre acesso aos dados de pesquisa, isto é, atuando na curadoria das informações produzidas e registradas por parte dos pesquisadores na condução da pesquisa científica. Dentre os estudos observados na literatura nacional, a proposta conduzida pelos referidos autores é a que mais se aproxima da perspectiva da Biblioteconomia de dados, ao passo que evidenciam a ação bibliotecária no processo. Nessa mesma perspectiva, a pesquisa apresentada por Semeler e Pinto (2019) trabalha os diferentes conceitos de dados de pesquisa como forma de demonstrar a emergência de uma Biblioteconomia orientada ao uso de dados, recorrendo sobre o importante papel de bibliotecários e cientistas da informação nesta nova abordagem.

No cenário internacional, destacam-se os estudos de Rousidis et al. (2014), Palavitsinis, Manouselis e Sanchez-Alonso (2014) e Barton, Currier e Hey (2013), que trazem eminentes preocupações acerca da qualidade dos metadados utilizados na descrição de conjuntos de dados armazenados em repositórios. Entretanto, é na pesquisa de Rice e Southall (2016) que

o presente trabalho encontra contribuições à questão aqui explorada.

Na obra *The Data Librarian's Handbook*, publicada em 2016, os autores compreendem a Biblioteconomia de dados como sendo a combinação das habilidades tradicionais da área, auxiliando os pesquisadores na descoberta de recursos informacionais, de um lado, e fornecendo informações em questões específicas de formatos de dados ou obsolescência digital, do outro (RICE; SOUTHALL, 2016, p. 16). Para os autores, essa nova concepção demanda que o profissional bibliotecário tenha conhecimento especializado e detalhado acerca das coleções, bem como seja capaz de orientar usuários sobre questões que envolvem o direito autoral, gerenciamento de dados, ação integrante do processo da pesquisa científica, assim como o uso de ferramentas de análise de dados mais apropriadas para cada público atendido.

Relativo aos usuários de dados, se inicialmente os serviços biblioteconômicos se pautavam em *meios físicos* e métodos de *entrega*, com o desenvolvimento das atuais tecnologias o tema dominante se tornou *acesso*. De modo complementar, defendem que, além da expertise, o bibliotecário de dados precisa ter familiaridade com a gama de problemas do campo da comunicação científica, considerando-se que a demanda por apoio à pesquisa vai além da descoberta de informações e acesso às mesmas, exigindo um envolvimento mais profundo com o processo de pesquisa e seus pormenores, como a gestão dos dados de pesquisa (RICE; SOUTHALL, 2016, p. 16).

Sobre isso, cabe o registro de um trecho da obra aqui trabalhada, a qual aponta a semelhança do trabalho do bibliotecário de dados com o realizado pelos bibliotecários acadêmicos tradicionais, a saber:

O trabalho diário do bibliotecário de dados [...] pode envolver trabalhar dentro de sistemas de bibliotecas, adquirindo recursos e desenvolvendo relações de trabalho que permitam promover o papel de sua biblioteca. O fato de trabalharmos com dados de pesquisa ao lado de periódicos, livros e outras publicações não devem fazer muita diferença de como o nosso trabalho é visto, mas existe uma série de razões que o faz. A palavra 'dados' em si é desanimadora para alguns bibliotecários e pesquisadores acadêmicos tradicionais e causa alguma ansiedade. Para alguns, é porque parece pertencer a outras disciplinas e ter pouca relação com seu próprio trabalho. Outros a veem como sendo uma palavra tão comum a ponto de ser quase indistinguível da 'informação' (RICE; SOUTHALL, 2016, p. 19, tradução nossa).

Nesse contexto, destacam-se os desafios que se colocam para a Organização do Conhecimento relacionados à descrição dos *metadados de dados de pesquisa* – tanto descritivos como temáticos – na condição de novos recursos informacionais e como estes se relacionam com os demais documentos do repositório de dados de pesquisa. Considerando este

cenário como ponto de partida, surge, então, um novo papel para a atuação do bibliotecário: o *bibliotecário de dados*.

3 O papel do bibliotecário de dados em repositórios de dados de pesquisa

No domínio dos repositórios de dados de pesquisa de bibliotecas acadêmicas, o bibliotecário de dados desempenha um papel-chave em dar suporte aos pesquisadores em todos os aspectos que cercam essa determinada classe de informação digital (uso, preservação e curadoria) e na produção e utilização de outros dados de pesquisa pelos pesquisadores em volumes cada vez maiores, visando à criação de novos conhecimentos (RICE; SOUTHALL, 2016, prefácio).

Aqueles que trabalham dentro de bibliotecas fazem uma valiosa contribuição para apoiar a pesquisa e o ensino, bem como moldar o caráter e a vida intelectual de instituições individuais. Quer o foco de uma universidade esteja nas humanidades, nas ciências físicas, nos clássicos ou em qualquer outro número de disciplinas, o bibliotecário, em última instância, trabalha para apoiar o aprendizado e a disseminação do conhecimento. Isso pode assumir muitas formas estabelecidas, mas cada vez mais existe a necessidade de apoiar novas formas de informação. Os dados digitais são uma nova forma. No caso de coleta de dados e criação de dados de pesquisa, isso também levou ao surgimento de um novo tipo de profissional de bibliotecas: o bibliotecário de dados. Mas até que ponto isso é, de fato, um novo papel e de que maneira ele difere da Biblioteconomia tradicional? [...] Seu papel é manipular, interpretar, analisar, observar ouvir ou, mais geralmente, “usar” os dados (RICE; SOUTHALL, 2016, p. 1-2, tradução nossa).

Rice e Southall (2016, p. 106) sinalizam que, ao incentivar o depósito dos dados de pesquisa no repositório de dados de pesquisa e responder com precisão às necessidades acadêmicas de pesquisadores relacionadas ao acesso, a manipulação e o compartilhamento de dados de pesquisa, o bibliotecário de dados cria um clima de confiança dentro da comunidade de pesquisa em relação ao repositório de dados de pesquisa.

Dentre as atribuições do bibliotecário de dados, destacam-se as questões-chave referentes à implantação e gestão dos repositórios de dados de pesquisa em bibliotecas acadêmicas. Nesse sentido, os autores consideram sete aspectos importantes na criação e gestão de repositórios de dados de pesquisa, quais sejam: definição do escopo; escolha do esquema de metadados; gerenciamento do acesso; revisão da qualidade dos dados; planejamento de preservação digital; promoção de repositórios digitais confiáveis; e necessidade de interoperabilidade.

A seguir, constam os princípios que regem cada um dos sete aspectos, de modo a contextualizar a sua inserção em repositórios de dados de pesquisa.

3.1 Definição do escopo

É essencial que o repositório de dados de pesquisa inclua os seguintes elementos: *assuntos e idiomas*; *tipo de dados de pesquisa* (por ex. observacional, experimental, computacional); *estado dos dados da pesquisa* (por ex. dados preliminares, apenas aqueles que fundamentam os resultados de pesquisa publicados ou apenas conjuntos de dados totalmente documentados); *versões*; *formatos de arquivo de dados*; e *limitações de volume e tamanho*. As decisões sobre a natureza dos dados a serem aceitos no repositório de dados de pesquisa, com base em critérios bem definidos, precisam ser realizadas com antecedência para evitar que o propósito de criação do repositório de dados de pesquisa seja comprometido. No caso de captura de dados multidisciplinares produzidos na universidade, o bibliotecário de dados precisa decidir os atores sociais que realizarão o depósito. Para tanto, um método de autenticação e autorização precisa ser adotado, tal como o uso de *login* único utilizado para outros serviços acadêmicos (RICE; SOUTHALL, 2016, p. 106-108).

3.2 Escolha do esquema de metadados

De acordo com Rice e Southall (2016, p. 24), os metadados também podem desempenhar um papel importante na gestão de dados de pesquisa. No contexto dos repositórios de dados de pesquisa, os metadados configuram-se como descrições técnicas altamente estruturadas que possibilitam documentar informações sobre como os dados foram criados, originalmente usados e o relacionamento entre os arquivos (por exemplo, entre um determinado artigo de periódico e o questionário utilizado na pesquisa). Nesse caso, metadados e documentação tornam-se sinônimos.

A utilização de um padrão de identificadores persistentes de objetos digitais como o DOI não apenas para os documentos científicos, mas também para os próprios conjuntos de dados de pesquisa, torna-se relevante por dois principais motivos: para que permaneçam válidos por mais tempo do que uma URL e possam ser remapeados caso o conteúdo seja movido entre *sites* ou em alguma rede desconhecida futuramente. Cada repositório de dados de pesquisa precisa equilibrar o número de campos fornecidos com a disposição dos depositantes ou recursos de equipe disponíveis para inserir metadados adicionais. Uma alternativa é o estabelecimento da obrigatoriedade de preenchimento de um número mínimo de campos e

simplesmente incentivar os depositantes a preencherem campos adicionais na medida do possível, visto que campos adicionais agregam valor ao registro de metadados e aprimoram a usabilidade dos dados. O esquema 4.0 do *DataCite* inclui os seguintes campos de metadados adicionais (recomendados e opcionais): *assunto, contribuidor, data, idioma, identificador alternativo, identificador relacionado, tamanho, formato, versão, direitos, descrição, localização geográfica, referência de financiamento* (RICE; SOUTHALL, 2016, p. 108-110).

3.3 Gerenciamento do acesso

O gerenciamento de acesso refere-se às decisões quanto aos tipos de acesso em um repositório de dados de pesquisa:

Acesso aberto: qualquer usuário com acesso à Internet pode acessar os dados de pesquisa, que podem ter termos de uso específicos ou indicar o uso apropriado ou impróprio através de uma licença padrão (como o pacote *Creative Commons*). Normalmente, é fornecido acesso anônimo aos dados, mas, em alguns casos, um nome e endereço de *e-mail* podem ser solicitados antes que o acesso seja concedido.

Acesso gerenciado: regras podem ser aplicadas ao uso dos dados (por ex. os usuários podem não apenas precisar se registrar, mas também ser aprovados antes que o acesso seja concedido). A aprovação pode depender do *status* do usuário (por ex. membro da instituição acadêmica) ou de suas respostas a determinadas perguntas (por ex. sobre seu objetivo da pesquisa).

Acesso seguro: os dados são liberados somente através de certos mecanismos seguros. Isso pode envolver o acesso a um servidor remoto para executar análises, em vez de baixar dados diretamente; ter a saída verificada pela equipe do repositório de dados de pesquisa para garantir a não divulgação de informações pessoais ou confidenciais; a realização de uma formação ou a análise detalhada dos procedimentos de análise, antes de o acesso ser concedido; deslocamento para um local especificado para acessar os dados em um computador sem rede; ou concordar com um contrato legal em que as ferramentas, regras ou padrões de manuseio e gerenciamento de dados sejam especificados e assinados pelo pesquisador e em um contato institucional.

Acesso fechado: um registro de metadados está disponível publicamente, porém os dados não estão prontamente acessíveis aos usuários. Pode haver um embargo de tempo limitado bloqueando o acesso ou o item pode ter sido retirado por algum motivo.

Tanto o bibliotecário de dados quanto os criadores/depositantes de dados de pesquisa devem

conhecer sobre os tipos de licenças existentes e a diferença entre acesso aberto “grátis” – qualquer pessoa pode ter uma cópia do item gratuitamente, mas quaisquer restrições de direitos autorais ou termos de licença permanecem válidos; e acesso aberto “livre” – além de ter uma cópia dos dados para uso próprio, é necessária uma permissão explícita para fazer cópias e divulgar os dados (normalmente por meio de licença aberta) (RICE; SOUTHALL, 2016, p. 111-112).

3.4 Revisão da qualidade dos dados

Considerando-se o desafio de avaliar o valor de um conjunto de dados ou o interesse futuro por esses dados ao longo do tempo, torna-se relevante a sistematização de verificações de controle de qualidade para cada conjunto de dados depositado no repositório de dados de pesquisa. Um ponto importante a ser verificado é que de forma geral os dados de pesquisa depositados em um repositório são as únicas cópias disponíveis, com exceção de algumas imagens, tabelas ou gráficos que aparecem no artigo publicado, tornando os dados de pesquisa adições únicas no repositório de dados de pesquisa. A revisão de qualidade pode, então, melhorar a utilidade real do material a ser preservado, bem como reforçar o compromisso da instituição em desenvolver e manter um repositório de dados de pesquisa. As principais decisões políticas em um repositório de dados de pesquisa se referem ao nível do controle de qualidade dos dados a ser aplicado na fase de depósito dos dados de pesquisa, à quantidade de curadoria e ao nível de compromisso com a preservação digital a longo prazo, que dependem da visão e missão do sistema, considerando as expectativas das partes interessadas (recursos humanos e financeiros da instituição). A avaliação do valor dos dados pelo bibliotecário de dados torna-se relevante na medida em que se um determinado conjunto de dados é visto como menos importante, é improvável que as técnicas e procedimentos especiais sejam desenvolvidos e aplicados. (RICE; SOUTHALL, 2016, p. 112-114).

Ainda para os autores, “qualquer controle de qualidade que possa ser aplicado no depósito (e mesmo antes), economizará tempo, esforço e frustração se os dados forem considerados merecedores de curadoria ou se um usuário solicitar auxílio no uso dos dados” (RICE; SOUTHALL, 2016, p. 113).

3.5 Planejamento de preservação digital

Diversos fatores contribuem para a fragilidade do material digital ao longo do tempo, que variam de problemas de *hardware* e *software* até mudanças culturais. Nesse sentido, a preservação digital é um requisito para evitar a deterioração do material ou para manter a continuidade de acesso para futuros usuários.

O bibliotecário de dados precisa lidar com a questão do planejamento de preservação digital ao longo do espaço e do tempo, visto que todas as ações de preservação digital envolvem recursos, tempo e esforço, e saber quais ações serão compensadas no futuro é realmente incerto. Na ótica dos autores, os principais problemas relacionados à preservação dos dados são:

“Bit rot”: deterioração ao longo do tempo de *bits* individuais que compõem arquivos, especialmente aqueles que não são processados e, portanto, copiados com muita frequência. Dependendo de quais *bits* realmente se danificam, pode tornar o arquivo ilegível ou pode fazer com que a informação desapareça em uma parte não percebida do arquivo;

Deterioração das próprias mídias: devem ser armazenadas de forma adequada, especialmente o armazenamento off-line, como CDs e DVDs, e atualizadas de acordo com um cronograma planejado para evitar a deterioração. Por essa razão, o armazenamento online (por ex. disco em servidores) pode ser preferível, mesmo para arquivos não acessados com muita frequência. Entretanto, de forma geral, tal alternativa demanda um alto custo.

Formatar obsolescência: atualmente, muitas bibliotecas precisam retirar os CD-ROMs não por sua deterioração física, mas porque o formato não é mais compatível com os sistemas operacionais e o *hardware* mais recentes.

Todavia, os autores alertam que nenhuma solução de armazenamento resolve completamente o problema da deterioração de *bits* ao longo do tempo; em vez disso, o problema precisa ser gerenciado pelo bibliotecário de dados. Entretanto, apontam três soluções indicadas por especialistas em preservação digital:

- 1) *Migração de formato*: compromisso de migrar ativamente os formatos para novas versões de *software* antes que eles corram o risco de se tornar obsoletos;
- 2) *Emulação*: engenharia de novos *softwares* que podem ler um formato obsoleto e processá-lo da mesma forma que o *software* original;
- 3) *Prescrição*: apenas aceitar formatos de arquivo que estejam em conformidade com os formatos de preservação padrão; uma variação é apenas aceitar arquivos que tenham formatos abertos. Os formatos de preservação padrão são amplamente conhecidos e compreendidos, seja por meio de uma especificação acordada por uma organização de padrões, como a ISO (*International Standards Organization*) ou devido ao uso de longo prazo (RICE; SOUTHALL, 2016, p.114-116).

3.6 Promoção de repositórios digitais confiáveis

Na visão dos autores, buscar o *status* de repositório digital confiável pode ajudar o repositório de dados de pesquisa a ganhar credibilidade e atingir os objetivos de planejamento de preservação digital. Tal *status* pode ser alcançado por meio do *The Data Seal of Approval* (DSA) – Selo de Aprovação de Dados. Os critérios para atribuir um DAS aos repositórios de dados de pesquisa obedecem às diretrizes nacionais e internacionais para arquivamento digital. Uma característica interessante do DAS é a divisão de responsabilidades, compreendendo três atores: o produtor de dados – correção futura de seus dados, bem como torná-lo compreensível para pesquisadores que talvez não estejam trabalhando no mesmo campo ou tenham acesso aos mesmos recursos de *software*; o consumidor de dados (usuário) – garantia de que os dados tenham qualidade suficiente para serem úteis em sua própria pesquisa; e o repositório de dados de pesquisa – gerenciamento, organização dos dados e interação com produtores de dados e consumidores de dados para auxiliá-los a fazer o melhor uso dos mesmos (RICE; SOUTHALL, 2016, p. 116-117).

3.7 Necessidade de interoperabilidade

A questão da interoperabilidade é um dos pontos mais importantes que o bibliotecário de dados precisa considerar em um repositório de dados de pesquisa. Na visão dos autores,

[...] é importante que o repositório de dados seja tão interoperável quanto possível. A última coisa que o mundo precisa é de outro portal de dados que acrescente valor e dê contexto aos dados, mas esteja desconectado da internet ou inacessível (RICE; SOUTHALL, 2016, p. 117, tradução nossa).

Para promover a interoperabilidade dos repositórios de dados de pesquisa, o uso de API (Interface de Programação de Aplicações) torna-se viável ao permitir que um sistema obtenha informações e passe informações para outros sistemas através de um conjunto de instruções e protocolos de programação. As formas baseadas em padrões de obtenção de informações equivalentes podem incluir o OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*), padrão específico de repositório que se baseia no Dublin Core para permitir a coleta de recursos (por ex. para um serviço ou portal agregado).

Outras maneiras importantes para os repositórios de dados de pesquisa serem interconectados: realizar o registro de repositório de dados de pesquisa no Re3Data.Org² para maior visibilidade; utilizar um serviço de identificador persistente (PID), como DataCite, para que o repositório seja encontrado por meio de uma pesquisa dos identificadores persistentes

de seus itens; ser coletado pelo *Data Citation Index* (Thomson Reuters), que coleta conteúdo de repositórios de dados de pesquisa selecionados para rastrear citações de conjuntos de dados na literatura; utilizar sistemas de autoridade de nomes internacionais, como o ORCID (*Open Researcher and Contributor ID*), que ajudará a vincular criadores de dados a outros produtos que produzem e a seus colaboradores.

Além de ser interoperável, um repositório confiável deve ser FAIR (*Findable, Accessible, Interoperable e Re-Usable*). A iniciativa *Data FAIRport* visa agregar valor aos dados científicos, possibilitando sua reutilização, garantindo metadados e outras documentações adequadas e capacidade de rastreamento de versões de origem e de dados. Aqui, o bibliotecário de dados avalia a necessidade de cada um desses pontos e o que é possível em seu ambiente institucional (RICE; SOUTHALL, 2016, p. 117-119).

4 Apontamentos finais

As discussões aqui conduzidas indicam que as questões técnicas sobre os repositórios de dados de pesquisa precisam ser desenvolvidas, gerenciadas e promovidas pelo bibliotecário de dados de modo a envolver não apenas a biblioteca universitária, mas toda a comunidade acadêmica. Nesse cenário, as bibliotecas universitárias estão em boa posição para criar e gerenciar repositórios institucionais e/ou de dados, pois além de estarem próximas dos pesquisadores que produzem os dados, são instituições de vida longa criadas para preservar e disseminar os registros acadêmicos, cada vez mais relevantes.

Embora as bibliotecas universitárias tenham uma expertise tradicional na descrição de recursos informacionais impressos, a descrição de objetos digitais e mais recentemente dos conjuntos de dados de pesquisa em repositórios de dados de pesquisa, configuram-se como um dos mais recentes e complexos desafios que se coloca à Organização do Conhecimento.

Cabe ressaltar que, mesmo que um repositório institucional de bibliotecas universitárias incorpore também os dados de pesquisa, continuará sendo denominado de repositório institucional e não repositório de dados de pesquisa, em atenção ao seu objetivo fulcral de reunir e possibilitar o acesso, uso e reuso das produções científicas, acadêmicas, artísticas e técnicas de uma determinada comunidade acadêmica que juntas conceituam o termo *produção intelectual*.

A atuação do bibliotecário universitário como bibliotecário de dados em repositórios de dados de pesquisa está diretamente relacionada à sua formação e educação continuada para lidar com os diversos aspectos que cercam a criação e gerenciamento de

metadados de dados como recursos informacionais digitais, bem como em todos os aspectos que cercam a produção e comunicação científica.

Diante do exposto, conclui-se que produtos e serviços relacionados ao uso e consumo de dados de pesquisa trazem uma perspectiva inovadora à atuação bibliotecária e às bibliotecas universitárias e serviços de informação científica. Dada a sua especificidade, os repositórios de dados de pesquisa exigem a observação de um conjunto de aspectos relacionados à cobertura do sistema, metadados, acesso e qualidade dos dados, preservação digital, confiabilidade dos sistemas e interoperabilidade. Esses princípios condicionam sobremaneira a condução de pesquisas que deem sustentação à gestão digital de dados de pesquisa e considerem as perspectivas e desafios que se apresentam para a organização da informação. Nesse viés, a observância da qualidade dos metadados utilizados na descrição de dados de pesquisa é necessária para a garantia do acesso. Por conseguinte, investigações complementares a esta pesquisa serão conduzidas de modo a discutir a representação e recuperação por assuntos em repositórios de dados de pesquisa e as especificidades a serem consideradas e discutidas nesta abordagem, de modo a avançar nas discussões aqui apresentadas.

Notas

- [1] The University of Edinburgh. *Information Services: Research Data Service*. Disponível em: <<http://www.ed.ac.uk/information-services/research-support/data-library/data-repository/definitions>>. Acesso em: 27 nov. 2018.
- [2] Criado em 2012 com o objetivo de promover uma cultura de compartilhamento, maior acesso e melhor visibilidade dos dados da pesquisa, trata-se de um registro internacional de repositórios de dados de pesquisa, que abrange diferentes áreas do conhecimento. Atualmente é financiado pela Fundação Alemã de Pesquisa Científica (DFG). Disponível em: <https://www.re3data.org>.

Referências

- AUCKLAND, M. Re-skilling for research: an investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers. *Research Libraries UK*, January, 2012.
- BERTIN, P. R. B.; VISOLI, M. C.; DRUCKER, D. P. A gestão de dados de pesquisa no contexto da e-science: benefícios, desafios e oportunidades para organizações de p&d. *Ponto de Acesso*, v. 11, n. 2, p. 34-48, 2017.
- CAVALCANTI, M. T.; SALES, L. F. Gestão de dados de pesquisa: um panorama da atuação da união

européia. *BIBLOS - Revista do Instituto de Ciências Humanas e da Informação*, v. 31, n. 1, p. 73-98, 2017.

CHOWDHURY, G. G.; CHOWDHURY, S. Information organization: issues and trends. In: _____. *Organizing information: from the shelf to the web*. London: Facet Publishing, 2007.

PALAVITSINIS, N.; MANOUSELIS, N.; SANCHEZ-ALONSO, S. Metadata quality in digital repositories: empirical results from the cross-domain transfer of a quality assurance process, *Journal of the Association of Information Science and Technology*, v. 65, n. 6, p. 1202–1216, 2014.

RICE, R.; SOUTHALL, J. *The data librarian's handbook*. London: Facet Publishing, 2016.

ROCHA, L. L.; SALES, L. F.; SAYÃO, L. F. Uso de cadernos eletrônicos de laboratório para as práticas de ciência aberta e preservação de dados de pesquisa. *Ponto de Acesso*, v. 11, n. 3, p. 2-16, 2017.

ROUSIDIS, D. et al. Metadata for Big Data: a preliminary investigation of metadata quality issues in research data repositories. *Information Services & Use*, n. 34, p. 279-286, 2014.

RUBIN, R. E. *Foundations of Library and Information Science*. 4th ed. London: Facet Publishing, 2016.

SANTOS, D. B.; ROCKEMBACH, M. Publicações ampliadas: aspectos da integração de dados de pesquisa. *Informação & Sociedade: Estudos*, v. 28, n. 2, 2018.

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

SEMELE, Alexandre Ribas; PINTO, Adilson Luiz. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. *Ciência da Informação*, v. 48, n. 1, p. 113-129, 2019.