

BIBLIOTECONOMIA DE DADOS EM REPOSITÓRIOS DE PESQUISA: PERSPECTIVAS PARA A ATUAÇÃO BIBLIOTECÁRIA

DATA LIBRARIANSHIP IN RESEARCH DATA REPOSITORIES: PERSPECTIVES ON THE LIBRARIAN'S ROLE

Roberta Cristina Dal'Evedove Tartarotti^a

Paula Regina Dal'Evedove^b

Mariângela Spotti Lopes Fujita^c

RESUMO

Introdução: 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. Estas novas concepções atrelam-se 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 desafiador.

Objetivo: Diante desse cenário, objetivou-se realizar pesquisa exploratória sobre os aspectos relacionados a esses novos conceitos no contexto das bibliotecas acadêmicas e a perspectiva que se apresenta para os estudos de organização da informação.

Metodologia: Pesquisa exploratória e qualitativa. **Resultados:** Observa-se que os repositórios de dados de pesquisa trazem uma perspectiva inovadora à atuação do bibliotecário ao demandarem questões relacionadas à cobertura do sistema, metadados, acesso e qualidade dos dados, preservação digital, confiabilidade dos sistemas e interoperabilidade. **Conclusões:** As questões relativas à Biblioteconomia de dados exigem o envolvimento de pesquisadores e profissionais na condução de pesquisas e ações dedicadas à gestão digital de dados de pesquisa e as perspectivas e desafios que se apresentam para a organização da informação.

Descritores: Dados de pesquisa. Repositórios de dados. Atuação bibliotecária.

^a Doutora em Ciência da Informação pela Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Coordenadora de Serviço da Diretoria de Difusão da Informação da Biblioteca Central César Lattes - Universidade Estadual de Campinas (UNICAMP). E-mail: roberta_tartarotti@yahoo.com.br.

^b Doutora em Ciência da Informação pela Universidade Estadual Paulista (UNESP). Docente do Departamento de Ciência da Informação da Universidade Federal de São Carlos (UFSCar). E-mail: p.dallevedove@gmail.com.

^c Doutora em Ciências da Comunicação pela Universidade de São Paulo (USP). Docente do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP). E-mail: mariangelaslf57@gmail.com

Biblioteconomia de dados. Bibliotecário de dados.

1 INTRODUÇÃO

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 pesquisa 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 o *conjunto de dados de pesquisa* decorrentes de um projeto de pesquisa, descritos e disponibilizados para uso da comunidade de pesquisadores cada vez mais diversificada, que não se limita à descoberta de informações e acesso, exigindo um envolvimento mais profundo com o processo de pesquisa (RICE; SOUTHALL, 2016).

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 de dados de pesquisa surgem para apoiar a produção, armazenamento, uso, reuso, acesso e compartilhamento dos conjuntos de dados gerados durante as várias etapas do processo da pesquisa científica.

Os dados de pesquisa surgem neste cenário de revolução como recursos essenciais e estratégicos na produção científica, sendo requeridos métodos que viabilizem o compartilhamento e o reuso 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 pesquisa”¹, tais como documentos, transcrições de entrevistas ou protocolos verbais, questionários, fotografias, vídeos,

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.

protocolos experimentais, cadernos de laboratório, diários de campo, planilhas, slides, amostras, gráficos, softwares 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. 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. Este 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 acadêmica, 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, em decorrência do avanço tecnológico e de uma sociedade cada vez mais imersa no ciberespaço, surgem modernos e inovadores sistemas de recuperação de informação. 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).

Em 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.

Isto posto, o trabalho apresenta resultados de uma pesquisa exploratória e qualitativa na literatura nacional e internacional sobre os aspectos relacionados aos conceitos Biblioteconomia de dados e bibliotecário de dados no contexto das bibliotecas acadêmicas e a 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: UM (NOVO) CAMPO DE ATUAÇÃO DO BIBLIOTECÁRIO

O conjunto das atividades de uma biblioteca acadêmica 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 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), Cavalcanti, Sales e Pimenta (2018), Santos e Rockembach (2018), Sayão e Sales (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, discorrendo sobre o importante papel de bibliotecários e cientistas da informação nesta nova abordagem.

Frente ao protagonismo dos dados de pesquisa, Sayão e Sales (2019) apresentam de modo exaustivo os elementos essenciais para a composição de um modelo de avaliação de sistemas de informação acadêmica voltados à gestão de dados de pesquisa, em atenção aos parâmetros técnicos, gerenciais e organizacionais requeridos. Para tanto, é preciso que os repositórios de dados sejam ambientes multifacetados, com condições bem definidas para uma gestão de dados compatível com as complexidades impostas pelos dados de pesquisa e, idealmente, disponham de serviços que os tornem espaços de colaboração e interação entre a comunidade de pesquisadores (SAYÃO; SALES, 2019).

No cenário internacional, destacam-se os estudos de Currier e Hey (2013), Rousidis, Garoufallou, Balatsoukas e Sicilia (2014), Palavitsinis, Manouselis e Sanchez-Alonso (2014), que trazem eminentes preocupações acerca da qualidade dos metadados utilizados na descrição de conjuntos de dados armazenados em repositórios. Como desdobramento dessas preocupações, Eynden e Corti (2017) incluem a contribuição de iniciativas a favor

de uma gestão de dados responsiva, que oriente pesquisadores em boas práticas na criação de dados compartilháveis, visando elevar o padrão dos repositórios de dados de pesquisa.

A gestão de dados de pesquisa é hoje um dos principais recursos para o avanço científico. Embora muitos pesquisadores reconheçam os benefícios do gerenciamento de dados para o compartilhamento de dados de pesquisa confiáveis, detectáveis, acessíveis e reutilizáveis, a falta de habilidades e conhecimentos técnicos, aliada à ausência de suporte institucional têm sido uma das principais barreiras a serem enfrentadas atualmente (AYDINOGLU; DOGAN; TASKIN, 2017).

Na obra *The Data Librarian's Handbook* publicada em 2016, Rice e Southall discutem o conceito de Biblioteconomia de dados e defendem que esta nova configuração da área alia habilidades tradicionais, auxiliando pesquisadores na descoberta de recursos informacionais, ao mesmo tempo que fornece informações em questões específicas de formatos de dados ou obsolescência digital. Esta 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 envolvam direito autoral, gerenciamento de dados, ação integrante do processo da pesquisa científica e uso de ferramentas de análise de dados mais apropriadas para cada público atendido (RICE; SOUTHALL, 2016).

Relativo aos usuários de dados, se inicialmente os serviços biblioteconômicos pautavam-se em *meios físicos* e métodos de *entrega*, com o desenvolvimento das atuais tecnologias o tema dominante tornou-se *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).

O trabalho do bibliotecário de dados em muito se assemelha aos

realizados pelos bibliotecários acadêmicos tradicionais, sendo que:

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).

Neste 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 esses 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.

3 O (NOVO) PAPEL DO BIBLIOTECÁRIO DE DADOS EM REPOSITÓRIOS DE DADOS DE PESQUISA

Como protagonistas ao aprendizado e à disseminação do conhecimento nas instituições acadêmicas, o bibliotecário favorece a integração com novas formas de informação. Os dados digitais fazem parte dessa nova configuração, a qual passou a integrar mais amplamente o ambiente das bibliotecas acadêmicas a partir da coleta de dados e criação de dados de pesquisa. Por decorrência, um novo perfil de profissional de bibliotecas foi assumido: o bibliotecário de dados. Todavia, os referidos autores questionam se de fato trata-se de um novo papel assumido pelo bibliotecário e de que maneira ele difere das atividades biblioteconômicas tradicionais.

No domínio dos repositórios de dados de pesquisa de bibliotecas acadêmicas, o bibliotecário de dados desempenha um papel-chave ao “[...] manipular, interpretar, analisar, observar ouvir ou, mais geralmente, ‘usar’ os dados” (RICE; SOUTHALL, 2016, p. 1-2, tradução nossa). Ao dar suporte aos pesquisadores em todos os aspectos que cercam esta 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, suas ações contribuem para a criação de novos conhecimentos (RICE; SOUTHALL, 2016).

Rice e Southall (2016, p. 106) sinalizam que ao incentivar o depósito dos dados de pesquisa e responder com precisão às necessidades acadêmicas de pesquisadores relacionadas ao acesso, à manipulação e ao compartilhamento de dados de pesquisa, o bibliotecário de dados cria um clima de confiança dentro da comunidade de pesquisadores em relação ao repositório de dados de pesquisa.

Dentre as atribuições do bibliotecário de dados, destaca-se a definição de subsídios que favoreçam a implantação e gestão dos repositórios de dados de pesquisa em bibliotecas acadêmicas. Nesse sentido, os referidos autores apresentam sete aspectos importantes a serem considerados na criação e gestão de repositórios de dados de pesquisa pelo bibliotecário de dados, quais sejam: definindo o escopo; escolhendo o esquema de metadados; gerenciando o acesso; revisando a qualidade dos dados; planejando a preservação digital; promovendo repositórios digitais confiáveis; e possibilitando a interoperabilidade.

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

3.1 DEFININDO O 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 ESCOLHENDO O ESQUEMA DE METADADOS

De acordo com Rice e Southal (2016, p. 24) os metadados também podem desempenhar um importante papel na gestão de dados de pesquisa pelo bibliotecário de dados. 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 (Digital Object Identifier) não apenas para os documentos científicos, mas 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, incentivando os depositantes a preencherem campos adicionais na medida do possível, visto que 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 GERENCIANDO O ACESSO

O gerenciamento de acesso contempla as decisões a serem tomadas pelo bibliotecário de dados 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 serem 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 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;

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

Tanto o bibliotecário de dados quanto os criadores/depositantes de dados de pesquisa devem conhecer os tipos de licença 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 REVISANDO A QUALIDADE DOS DADOS

Considerando-se o desafio do bibliotecário de dados em avaliar o valor de um conjunto de dados ou o interesse futuro por estes 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 referem-se ao nível do controle da 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 é estimulada na medida em que se um determinado conjunto de dados é visto como menos relevante, 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 antes ou no momento do depósito dos dados em um repositório de dados de pesquisa economizará tempo, esforço e frustração se os dados forem considerados merecedores de curadoria (RICE; SOUTHALL, 2016, p. 113).

3.5 PLANEJANDO A 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 destas ações serão compensadas futuramente é incerto. Na ótica de Rice e Southall (2016), 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 fazer com que a informação desapareça em uma parte não percebida do arquivo;

- *Deterioração das próprias mídias*: as 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 esta razão, o armazenamento online pode ser preferível, mesmo para arquivos não acessados com muita frequência. Por outro lado, tal alternativa demanda um alto custo.

- *Formatar obsolescência*: atualmente, muitas bibliotecas acadêmicas precisam retirar os CD-ROMs não por sua deterioração física, mas porque o formato não é mais compatível com os sistemas operacionais e os hardwares 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. Como alternativa, são propostas três soluções indicadas por especialistas em preservação digital:

- *Migração de formato*: compromisso de migrar ativamente os formatos para novas versões de software antes que corram o risco de se tornarem obsoletos;

- *Emulação*: engenharia de novos softwares que podem propiciar a leitura de um formato obsoleto e processá-lo da mesma forma que o software original;

- *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 PROMOVENDO REPOSITÓRIOS DIGITAIS CONFIÁVEIS

Rice e Southall (2016) consideram que a busca pelo 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 DSA aos repositórios de dados de pesquisa obedecem às diretrizes nacionais e internacionais para arquivamento digital. Uma característica interessante do DSA 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 do bibliotecário de dados com produtores de dados e consumidores de dados para auxiliá-los, de modo que façam o melhor uso possível (RICE; SOUTHALL, 2016, p. 116-117).

3.7 POSSIBILITANDO A 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. Neste entendimento, ser interoperável é um dos requisitos essenciais, ao passo que “[...] 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 formas importantes de viabilizar o diálogo entre os sistemas de gestão de dados de pesquisa são: realizar o registro do 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 de dados confiável deve ser FAIR (*Findable, Accesible, 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 destes pontos e o que é possível em seu ambiente institucional (RICE; SOUTHALL, 2016, p. 117-119).

4 CONSIDERAÇÕES FINAIS

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>.

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 em si, mas toda a comunidade acadêmica. Neste cenário, as bibliotecas acadêmicas estão em boa posição para criar e gerenciar repositórios de dados de pesquisa, 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.

A atuação do bibliotecário acadêmico 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 de pesquisa como recursos informacionais digitais, bem como em todos os aspectos que cercam a produção e comunicação científica.

Diante o 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 acadêmicas 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.

Embora as bibliotecas acadêmicas 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.

Neste viés, a observância da qualidade dos metadados utilizados na descrição de dados de pesquisa é necessária para a garantia do acesso. Por

consequente, 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 questões aqui apresentadas.

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.

AYDINOGLU, A.; DOGAN, G.; TASKIN, Z. Research data management in Turkey: perceptions and practices. **Library Hi Tech**, v. 35, n. 2, p. 271-289, 2017.

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

CAVALCANTI, M. T.; SALES, L. F.; PIMENTA, R. M. Compartilhamento e acesso a dados de pesquisa em humanidades digitais. **Cadernos BAD (Portugual)**, n. 1, p. 30-40, 2018.

CHOWDHURY, G. G.; CHOWDHURY, S. **Organizing information: from the shelf to the web**. London: Facet Publishing, 2007.

EYDEN, V.; CORTI, L. Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. **International Journal on Digital Libraries**, v. 18, n. 2, p. 113-121.

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.; GAROUFALLOU, E.; BALATSOUKAS, P.; SICILIA, M.-A. 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.

SAYÃO, L. F.; SALES, L. F. Subsídios para a construção de um modelo de avaliação de sistemas de gestão de dados de pesquisa. **Ponto de Acesso**, v. 12, n. 3, p. 80-108, 2018.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia e dados. **Ciência da Informação**, v. 48, n. 1, p. 113-129, 2019.

DATA LIBRARIANSHIP IN RESEARCH DATA REPOSITORIES: PERSPECTIVES ON THE LIBRARIAN'S ROLE

ABSTRACT

Introduction: 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. **Objective:** In view of this scenario, the objective was to conduct exploratory research on the aspects related to these new concepts in the context of academic libraries and the perspective that is presented for studies of information organization. **Methodology:** Exploratory and qualitative research. **Results:** It is observed that research data repositories bring and innovative perspective to the librarian's role, demanding issues related to system coverage, metadata, data access and quality, digital preservation, system reliability and interoperability. **Conclusions:** Questions concerning the Data librarianship requires 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.

Descriptors: Research data. Data repository. Librarian's role. Data librarianship. Data librarian.

BIBLIOTECOLOGÍA DE DATOS EN REPOSITARIOS DE DATOS DE INVESTIGACIÓN: PERSPECTIVAS PARA LA ACTUACIÓN BIBLIOTECARIA

RESUMEN

Introducción: Los repositorios de datos de investigación emergen como sistemas contemporáneos y heterogéneos para recuperar información científica, en los que se destacan dos conceptos: Bibliotecología de datos y Bibliotecario de datos. Estos nuevos conceptos están vinculados a discusiones en la Organización del Conocimiento en defensa de instituciones efectivas y sostenibles, en atención a los métodos y prácticas empleados para el análisis de datos de investigación, especialmente en el entorno web, haciendo mucho el papel del bibliotecario en la organización de estos sistemas más desafiante. **Objetivo:** Dado este escenario, el objetivo fue realizar una investigación exploratoria sobre los aspectos relacionados con estos nuevos conceptos en el contexto de las bibliotecas académicas y la perspectiva presentada para los estudios de organización de la información. **Metodología:** Investigación exploratoria y cualitativa. **Resultados:** Se observa que los repositorios de datos de investigación aportan una perspectiva innovadora al desempeño del bibliotecario al exigir problemas relacionados con la cobertura del sistema, los metadatos, el acceso y la calidad de los datos, la preservación digital, la confiabilidad e interoperabilidad de los sistemas. **Conclusiones:** Los problemas de la Bibliotecología de datos requieren la participación de investigadores y profesionales en la realización de investigaciones y acciones dedicadas a la gestión digital de datos de investigación y las perspectivas y desafíos presentados para la organización de la información.

Descriptores: Datos de investigación. Repositorios de datos. Actuación bibliotecaria. Bibliotecología de datos. Bibliotecario de datos.