



REPRESENTAÇÃO DOCUMENTAL E DIGITALIDADE

percepções relacionais
entre tradição e inovação

Abecin
EDITORA

Paula Regina Dal'Evedove
Zaira Regina Zafalon
(organizadoras)

**REPRESENTAÇÃO DOCUMENTAL E
DIGITALIDADE**
**percepções relacionais entre tradição e
inovação**

São Paulo
Abecin Editora
2024

©2024 by Paula Regina Dal'Evedove e Zaira Regina Zafalon (organizadoras)
Direitos desta edição reservados à ABECIN Editora

Rua Maracaju, nº 58, Vila Mariana, São Paulo/SP, CEP 04013-020

ESSA OBRA É LICENCIADA POR UMA

LICENÇA CREATIVE COMMONS



Atribuição – Uso Não Comercial – Compartilhamento pela mesma licença 3.0

É permitido copiar, distribuir, exibir, executar a obra e criar obras derivadas desde que sem fins comerciais e que seja dado o crédito apropriado aos autores e compartilhada sob a mesma licença do original.

Dados Internacionais de Catalogação-na-Publicação (CIP)

R425 Representação documental e digitalidade: percepções relacionais entre tradição e inovação / Paula Regina Dal'Evedove e Zaira Regina Zafalon (organizadoras). – São Paulo: Abecin Editora, 2024.
259 p.

e-ISBN: 978-65-86228-13-7.

Disponível em: <https://portal.abecin.org.br/editora>.

1. Representação documental. 2. Tratamento da informação. 3. Ambientes informacionais digitais. I. Dal'Evedove, Paula Regina, org. II. Zafalon, Zaira Regina, org.

CDU: 02(05)

CDD: 020

Ficha catalográfica: Cleide Oliveira da Silva – CRB-8 8157/SP.

COMISSÃO EDITORIAL E CIENTÍFICA

Editor-chefe: Zaira Regina Zafalon (UFSCar)

Aldinar Martins Bottentuit (UFMA)	José Antonio Frías (USAL, Espanha)
Alessandra dos Santos Araújo (UFS)	José Antonio Moreiro González (UC3M, Espanha)
Andréa Pereira dos Santos (UFG)	Manuela Moro Cabero (USAL, Espanha)
Aurora Cuevas-Cerveró (UCM, Espanha)	Márcia Ivo Braz (UFPE)
Célia Regina Simonetti Barbalho (UFAM)	Márcio Bezerra da Silva (UNB)
Danielly Oliveira Inomata (UFAM)	Marta Lúgia Pomim Valentim (UNESP)
Dunia Llanes Padrón (UH, Cuba)	Martha Suzana Cabral Nunes (UFS)
Franciele Marques Redigolo (UFPA)	Meri Nadia Marques Gerlin (UFES)
Helen Beatriz Frota Rozados (UFRGS)	Naira Christofolletti Silveira (UNIRIO)
Henriette Ferreira Gomes (UFBA)	Paulina Szafran (UDELAR, Uruguai)
Ieda Pelógia Martins Damian (USP)	Samile Andréa de Souza Vanz (UFRGS)
Isidoro Gil Leiva (UM, Espanha)	Valéria Martin Valls (FESP/SP)
Ivana Lins (UFBA)	

Revisão: Paula Regina Dal'Evedove e Zaira Regina Zafalon
Diagramação e Editoração: Zaira Regina Zafalon

Normalização: Autores
Capa: Paula Regina Dal'Evedove

O conteúdo dos capítulos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores e não representam necessariamente a posição oficial da Editora Abecin. Os originais foram previamente submetidos à avaliação cega pelos pares, membros da Comissão Editorial e Científica desta Editora, tendo sido aprovados para a publicação.

CAPÍTULO 2

METADADOS PARA REPRESENTAÇÃO DE CONJUNTOS DE DADOS DE PESQUISA E OS PRINCÍPIOS FAIR

Caterina Marta Groposo Pavão

Cleusa Pavan

1 INTRODUÇÃO

O compartilhamento de dados de pesquisa tem crescido como efeito da demanda das agências de financiamento públicas e privadas e dos periódicos científicos, para os cientistas com projetos de pesquisa aprovados e no processo de *peer review*, respectivamente. Porém, mais do que cumprir com requerimentos, ao disponibilizar dados de pesquisa em acesso aberto, na internet, o cientista beneficia a sua área do conhecimento como um todo, encorajando os pares a realizarem novas investigações e descobertas. Sob o aspecto individual e de sua carreira, ele usufrui de uma plataforma própria, os repositórios, para organizar, preservar e disponibilizar dados de forma imediata, amplia a visibilidade de suas investigações e recebe crédito por meio da citação.

Esse novo processo está no escopo da Ciência Aberta que tem como princípio básico a convicção de que o conhecimento é um bem público e que grande parte das pesquisas científicas é desenvolvida com fomento governamental, devendo ficar mais acessível para a

comunidade científica, o setor de negócios e a sociedade em geral. Segundo o documento de recomendações da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) sobre Ciência Aberta, publicado em 2020, as soluções sustentáveis e inovadoras requerem esforços científicos eficientes, transparentes e dinâmicos, não somente da comunidade científica, mas também de toda a sociedade. Para assegurar que a ciência realmente favoreça as pessoas e o planeta, é necessário transformar todo o processo científico. Sendo assim, a Ciência Aberta é o movimento que pretende essa transformação, disponibilizando os dados, a informação e a produção científica ao alcance de todos, de forma democrática e confiável.

Esse movimento engloba uma série de práticas, além dos dados de investigação, como softwares e ferramentas de código aberto, fluxos de trabalho abertos, o acesso aberto às publicações, ciência cidadã, recursos educativos abertos, ciência aberta reprodutível, políticas para ciência aberta e métodos alternativos para avaliar a investigação, incluindo a revisão pelos pares aberta (Pontika *et al.*, 2015). Nesta estrutura da Ciência Aberta, os dados de pesquisa ocupam um lugar extremamente relevante no momento em que cada vez mais os recursos são escassos e partilhar conhecimento é a palavra de ordem. O compartilhamento de dados de pesquisa é basicamente o ato de disponibilizá-los para sua reutilização.

Os dados da pesquisa abrangem todas as disciplinas e, portanto, sua definição ou conceito pode variar em relação às abordagens dos diferentes atores (pesquisadores, instituições, órgãos de fomento, governos, entre outros) e dos diferentes

contextos nacionais. Resumidamente, podemos dizer que são as informações registradas ou produzidas por qualquer forma ou meio durante o processo de investigação. Em 2007, a *Organization for Economic Co-operation and Development* (OECD) publicou um conjunto de diretrizes com base em princípios comumente aceitos para facilitar o acesso otimizado e econômico aos dados de pesquisa digital financiados com recursos públicos,

No contexto destes princípios e diretrizes, os dados de pesquisa são definidos como registros factuais (registros numéricos, textuais, imagens e sons) usados como fontes primárias para pesquisa científica, e que são comumente aceitos na comunidade científica como necessários para validar resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto que está sendo investigado. (OECD, 2007, p. 13, tradução nossa)

Chapman *et al.* (2020) definem um conjunto de dados como um conjunto de imagens, gráficos ou documentos, bem como a tabela clássica de dados, relacionadas, organizadas e formatadas para um propósito específico. Os autores, ainda, colocam que a busca de conjuntos de dados envolve a descoberta, a exploração e o retorno de conjuntos de dados, como resultado, a um usuário final. Segundo Löffler *et al.* (2021), esta definição é um pouco restritiva, pois todos os tipos de dados científicos, como os experimentais, observacionais, ambientais e genômicos, simulações e dados computacionais podem ser considerados como conjuntos de dados. Assim, fazem uma releitura da definição de Chapman e colegas, da seguinte forma: “[...] um conjunto de dados é uma coleção de

dados científicos, incluindo dados primários e metadados organizados e formatados para um propósito específico” (Löffler *et al.*, 2021, p. 5, tradução nossa).

Ainda, conjuntos de dados de pesquisa podem ser considerados recursos com atributos que devem ser descritos por meio de estruturas de representação, que permitam sua busca e recuperação em um determinado sistema. A sistematização de dados de pesquisa é um processo que consiste em organizar os dados por meio de normas e padrões interoperáveis, que permitem sua recuperação e consequente acesso pela comunidade científica e acadêmica. Assim, os dados da pesquisa devem ser organizados para encorajar uma possível reutilização pelos sistemas abertos (Ávila Barrientos, 2020).

Os dados de pesquisa são o resultado dos projetos de investigação, usados como fontes primárias que sustentam a investigação científica e possibilitam a derivação de descobertas teóricas ou aplicadas. Devem, fundamentalmente, tornar as descobertas ou os estudos replicáveis, ou pelo menos reprodutíveis, ou precisam ser, em alguma medida, reutilizáveis. Para tal, Sales e Sayão (2019), colocam que os dados devem ser documentados, ter proveniência comprovada, estar armazenados em lugar seguro onde a preservação seja garantida para usos atuais e futuros, ser compartilhados, acessados e reusados. A fim de atender estas premissas, as boas práticas orientam para que os dados sejam o mais abertos e FAIR possíveis, mas sem esquecer as questões éticas, comerciais e de privacidade de dados sensíveis ou dados proprietários.

Os princípios FAIR surgiram em 2014, com o objetivo de apresentar normas e boas práticas para facilitar a reutilização dos dados de pesquisa. A sigla identifica as quatro características que os dados ou os objetos digitais devem possuir: *Findable* (localizáveis), *Accessible* (acessíveis), *Interoperable* (interoperáveis) e *Reusable* (reutilizáveis). De acordo com Rocha *et al.* (2019), para os dados serem qualificados como FAIR, a estes devem ser atribuídos identificadores únicos, persistentes e globais (**F**). Devem ser descritos por metadados indexáveis e ricos (**F**), representados em linguagens formais (**I**), aceitos pela comunidade (**R**), com atributos relevantes, precisos e úteis ao contexto (**R**), incluindo metadados de proveniência (**R**) e usando vocabulários controlados que seguem princípios FAIR (**I**). Estes dados devem ser recuperáveis pelo seu identificador, através de um protocolo de comunicação padronizado, aberto, gratuito (**A**). Também, devem ser acompanhados de licenças claras e acessíveis (**R**), e referências qualificadas, devem ligar (meta)dados para enriquecer o conhecimento sobre os mesmos (**I**). Os princípios FAIR não são um padrão, não ditam uma especificação tecnológica ou uma solução para implementação, mas um guia para assistir produtores e publicadores de dados em suas escolhas na gestão dos mesmos, enfatizando o aumento da capacidade das máquinas encontrarem e utilizarem automaticamente os dados, para além do reuso pelos indivíduos (Wilkinson *et al.*, 2016).

O objetivo deste capítulo é discorrer sobre a representação dos dados de pesquisa, com o intuito de torná-los localizáveis, acessíveis, interoperáveis e reutilizáveis.

Utilizar metadados completos e precisos é importante em vários processos, como o reuso e compartilhamento de conjuntos de dados entre os cientistas, a aplicação de curadoria digital e estratégias de proveniência dos dados, e a análise do conteúdo dos repositórios (Rousidis *et al.*, 2014). Os princípios FAIR são mínimos e devem ser observados especialmente pelos produtores de dados, que executam o autoarquivamento de conjuntos ou pacotes de dados (*datasets*, em inglês), e pelos mantenedores de repositórios. Neste capítulo, abordamos a representação da informação no contexto dos repositórios de dados. Para tal, trataremos questões relativas ao emprego de metadados para descrição e a sua qualidade, padrões de metadados internacionalmente aceitos pelas áreas do conhecimento, os valores contidos nos metadados, o cientista produtor, depositante e catalogador de dados, em repositórios de dados de pesquisa, e o papel do profissional da informação.

2 REPOSITÓRIOS DE DADOS DE PESQUISA

Um repositório é um serviço de informação dedicado ao gerenciamento, reunião, armazenamento, organização, preservação, recuperação e ampla disseminação da informação produzida em uma instituição por seus pesquisadores. Constitui-se em uma coleção de objetos digitais que consiste, basicamente, em um banco de dados, no qual se armazena o conteúdo em formato digital, em um servidor web, com um conjunto de interfaces e funções para envio de conteúdo, gerenciamento de sistema, pesquisa e recuperação de conteúdo e outros serviços para o usuário. Os

repositórios digitais devem refletir o conceito de "ambiente de confiança", que foi promovido durante a *International Conference 2012 Cultural Heritage online - Trusted Digital Repositories*, desenvolvido a partir de requisitos, normas, recomendações e diretrizes confiáveis e amplamente aceitas.

Um repositório digital, para ser considerado confiável, deve ter como missão fornecer acesso de longo prazo aos recursos digitais por ele gerenciados e, portanto, aceita a responsabilidade pela manutenção, também a longo prazo, desses recursos, em nome de seus depositantes e em benefício dos usuários atuais e futuros. Para que esse propósito seja atingido, devem seguir o *Open Archival Information System (OAIS)*, um padrão de referência para o desenvolvimento de repositórios que estão comprometidos em preservar e manter seus conteúdos digitais acessíveis a longo prazo. Ele especifica o ambiente do repositório, a forma de representação das informações a serem preservadas e o modelo funcional do repositório. Também, orienta na identificação dos produtores, do processo de produção dos dados, das comunidades-alvo (consumidores) e dos encarregados de sua gestão, especifica as entidades funcionais que compõem um repositório: ingestão, acesso, planejamento da preservação, armazenamento, gestão de dados e administração, serve para delimitar, caracterizar e orientar os componentes de funcionamento do repositório.

Os repositórios de dados exercem um papel fundamental na ciência. Implementam práticas sistemáticas de gestão, incluindo a curadoria, preservação, disponibilidade a longo prazo, disseminação e acesso. São amplamente

difundidos nas comunidades que os produzem, como na Física (SDSS - FermiLab, na Genética (GenBank na NCBI Data) ou nas Ciências ambientais (CEDA Archive).

Para cumprir tais papéis, os repositórios de dados, por meio da tecnologia, devem garantir a descoberta dos dados ao longo do tempo, mas para atender o quarto princípio FAIR, dependem, pelo menos parcialmente, do correto uso dos metadados e da quantidade dos mesmos, para que a descrição de dados reflita as necessidades de informações acadêmicas. Se a informação não está disponível nos metadados, os dados não podem ser recuperados, reutilizados e citados. O estudo de Löffler *et al.* (2021), na área de Biodiversidade, revela que grandes repositórios de dados seguem apenas parcialmente a recomendação do quarto princípio FAIR. Os repositórios de dados gerais tendem a usar apenas os metadados padrão, embora existam diretrizes e padrões de dados de metadados adequados. Assim, os autores se fazem os seguintes questionamentos: por que os padrões adequados às áreas, com as suas especificidades são tão raramente usados? As recomendações são muito amplas? Os padrões contêm muitos campos e elementos? Precisa-se melhorar as ferramentas de software e métodos para facilitar a integração de diferentes metadados? Os autores acreditam que apenas mais estudos de usuários e discussões na comunidade de pesquisa podem ajudar a encontrar respostas para essas perguntas.

Os repositórios de dados, segundo Monteiro e Sant'Ana (2018), têm sua gênese com a necessidade de gestão dos dados científicos, estão vinculados às universidades e instituições de pesquisa e contribuem para assegurar que os

dados sejam publicados e disponibilizados para a comunidade científica com o menor número possível de restrições. Podemos encontrar diversos desafios que tornam difícil descobrir, compartilhar ou reutilizar dados, como os metadados incompletos, imprecisos e/ou mal descritos. Outras barreiras são a baixa visibilidade, problemas com acessibilidade (acesso restrito, contato necessário com o produtor) e usabilidade limitada, devido aos metadados, como ausência de parâmetros-chave para avaliar a proveniência e qualidade dos dados, e para identificar a diversidade da amostra da população (quando for o caso) (Khan *et al.*, 2021). Os autores, que são vinculados às Ciências da Saúde, demonstram receio de os usuários secundários utilizarem *datasets* problemáticos, sem metadados básicos que informem aspectos importantíssimos na pesquisa clínica, sobre a população, como idade, sexo, raça, pacientes grávidas, além de amostras pequenas, pelo simples fato desses dados estarem disponíveis e terem alta visibilidade e fácil acesso no momento da busca em repositórios.

Para Ávila Barrientos (2020), o principal desafio é convencer as instituições a participarem do modelo de acesso aberto, implementando políticas de promoção de dados abertos. O estabelecimento de políticas institucionais norteia e incentiva os pesquisadores para uma melhor administração dos dados, inclusive dos conjuntos coletados há mais tempo e que ainda têm utilidade para a comunidade da área, podendo ser compartilhados tardiamente.

Com a consciência crescente em relação à citação de dados e que, para isso, os conjuntos de dados precisam ser

preservados, descritos e acessados corretamente, as atividades relacionadas à curadoria de dados tornam-se cada vez mais importantes. A fim de que um conjunto de dados seja citado, ele deve primeiro ser arquivado em um repositório, preservado em uma forma interoperável, adequadamente descrito por um conjunto formal de metadados conectados ao conjunto de dados e disponibilizados para reutilização (Corrêa, 2016). Vários padrões de metadados de metadados integral tal conjunto formal. Em alguns casos descrevem o conteúdo de arquivos, em outros as características técnicas dos arquivos e em outros expressam as relações entre os arquivos de um conjunto de dados. Ainda, alguns padrões de metadados servem como suporte para estabelecer uma estratégia de metadados capaz de descrever seus dados e atender às suas necessidades de gerenciamento.

3 METADADOS PARA REPRESENTAÇÃO DE DADOS DE PESQUISA

Na opinião de Solodovnik (2013), não importa a prática de gestão de informação escolhida pela instituição para processar dados bibliográficos, a qualidade dos mesmos depende da sua adequação aos requisitos das normas bibliográficas, bem como, os códigos de catalogação e regras de classificação e vocabulários controlados, utilizados não apenas para definir e descrever registros formalmente, mas também para normalizar seus dados, atributos e valores e, assim, criar um acesso controlado e uniforme de recuperação, interoperabilidade e reutilização de dados bibliográficos. Se bem esta citação refere-se a dados bibliográficos, os princípios

apresentados aplicam-se a outros tipos de conteúdo, documentos ou objetos, sejam eles digitais ou não.

A catalogação tradicional organiza todo um conjunto de dados: autor, título, data, assunto, entre outros, armazenados em um sistema que contém as coleções de uma biblioteca. Os metadados correspondentes a estes dados são considerados componentes lógicos centrais do sistema utilizado em bibliotecas e repositórios digitais. Segundo Campbell (2006), com o surgimento de novos sistemas de metadados, que se encontram nos próprios documentos e afetam a maneira como eles são usados, a catalogação mudou-se para um novo território difícil, em que nosso papel tradicional como intermediários de informação tornou-se menos rigidamente definido e mais sutilmente poderoso.

O esquema de metadados que se pretende utilizar em um repositório digital, seja de documentos ou de dados, institucional ou temático, deve determinar os requisitos de descrição de cada elemento, promover a padronização, normalização e enriquecimento dos metadados para fortalecer a qualidade dos registros e da informação contida nos objetos digitais. Antes de disponibilizar um objeto digital, deve-se ter certeza de que todos os metadados necessários foram registrados. Ainda, no momento do planejamento de um repositório, deve-se avaliar se um único esquema de metadados atende a natureza de todas as coleções ou se será necessário combinar esquemas para melhor atender às necessidades de descrição (Pavão *et al.*, 2015).

Da perspectiva dos dados, os metadados são informações estruturadas e descritivas, criados para

finalidades diferentes, como pesquisa, classificação ou formação de conhecimento e devem dar resposta para: o que foi medido/coletado/pesquisado? por quem? quando? onde? e por quê? Sales e Sayão (2019) argumentam que, para os dados serem encontrados, identificados, compreendidos, reusados e tenham sua proveniência comprovada, eles devem estar acompanhados de um conjunto de informações que descreva todos os seus aspectos, como ilustra a Figura 1.

Figura 1: Conteúdo da descrição dos dados



Fonte: Sales e Sayão (2019).

Esquemas ou padrões de metadados podem ser de amplo espectro, *Dublin Core* e *DataCite*, destacados abaixo, ou voltados a um domínio, como, por exemplo: *Darwin Core* (Biodiversidade), *Data Documentation Initiative* (Ciências sociais e econômicas), *Ecological Metadata Language* (Ecologia e Ciências ambientais e da terra), *EngMeta* (Engenharia computacional) e *ISO 19115* (Ciências geoespaciais).

Dublin Core Metadata Initiative: criado na década de 1990, é um padrão simples e flexível, de fácil uso por não catalogadores, com 15 elementos básicos para descrição de

recursos/objetos na internet, (Colaborador, Cobertura, Criador, Data, Descrição, Formato, Identificador, Língua, Publicador, Relação, Direitos, Recurso, Assunto, Título e Tipo).

DataCite Metadata Schema Documentation: esquema genérico para descoberta e citação de dados, desenvolvido pelo *DataCite Consortium*, mas que possibilita a customização para atender às disciplinas. Apresenta três níveis de obrigação: mandatório, recomendado e opcional. As propriedades obrigatórias são: Identificador, Criador, Título, Publicador, Ano de publicação, Tipo de recurso, enquanto Assunto, Colaborador, Data, Identificador relacionado, Descrição e Geolocalização são recomendadas (DATA CITE, 2021).

Para selecionar os elementos de metadados que serão utilizados, deve-se fazer o confronto contínuo com os seis princípios dos “bons metadados” publicados pela National Information Standards Organization (NISO) (2007), quais sejam:

- 1) devem estar em conformidade com os padrões da comunidade, de forma apropriada para os materiais da coleção, os usuários da coleção e para os atuais e potenciais futuros usuários;
- 2) oferecer suporte à interoperabilidade;
- 3) usar controle de autoridades e de conteúdo para descrever e relacionar objetos;
- 4) incluir uma declaração clara das condições e termos de uso para o objeto digital;
- 5) oferecer preservação e curadoria digital a longo prazo dos objetos nas coleções;

6) são objetos em si e, portanto, devem ter as qualidades de bons objetos, incluindo autoridade, autenticidade, arquivamento, persistência e identificador único.

Podemos verificar que os princípios FAIR apresentam características que se sobrepõem àqueles da NISO, pois consideraram os metadados um componente-chave para o compartilhamento de *datasets*, sendo valorados tanto quanto os dados em si. Somente uma descrição normalizada leva à recuperação e reuso dos conjuntos de dados num repositório de forma mais eficiente.

Para Solodovnik (2013), a validação da qualidade de metadados é uma questão de gestão organizacional em vez de um procedimento. Portanto, é uma boa prática estabelecer uma unidade de apoio dirigida por profissionais de metadados dentro de cada repositório. A redação de instruções e recomendações são essenciais para apoio ao produtor de dados.

Tais profissionais devem buscar o aprimoramento dos metadados de descrição de conteúdo, considerando a missão de manter um padrão de alta qualidade, em conformidade com as mais recentes tecnologias, e desenvolver recursos, a fim de que os conteúdos possam ser inequivocamente identificados, tendo em vista a sua disseminação. Wu *et al.* (2019) elegem dez recomendações para os repositórios aperfeiçoarem a descoberta de dados e a experiência dos usuários no momento da busca. Entre elas, tornar os registros de metadados legíveis e analisáveis, sendo que os campos mais importantes deveriam ficar no topo das páginas, nomeados

devidamente para não gerar ambiguidades, com *hyperlinks* e botões facilmente reconhecíveis; identificar e agregar registros de metadados de um mesmo *dataset*, uma vez que podem ser disponibilizados em outros repositórios, de forma diferente, quando fruto de pesquisa em colaboração entre instituições; e facilitar a indexação e busca dos registros de metadados pelos motores de busca de *datasets*, empregando ferramentas existentes para tal.

É preciso ressaltar que, em oposição aos catálogos de bibliotecas, bases de dados e outros serviços especializados de indexação, em que o profissional da informação é o principal responsável pela descrição dos objetos, nos repositórios de dados de pesquisa a atividade é majoritariamente do cientista. White (2014) argumenta que os dados de pesquisa são um tipo de informação pessoal, organizada de forma pessoal, fatos que precisam ser considerados antes de criar sistemas de informação e integrar qualquer um dos esquemas de metadados já estabelecidos. A coleta e organização de dados, para uso próprio e de seus colaboradores, é uma ação básica do fazer científico diário. Entretanto, organizar com a finalidade de compartilhar os dados, de forma ampla e global, somente tornou-se um requerimento, mais recentemente, para a maioria dos pesquisadores. Isso implica alterar os procedimentos, até então praticados, e dedicar mais tempo à descrição dos *datasets*, com ênfase na descoberta e compartilhamento e não apenas na análise para publicação nos tradicionais canais de comunicação científica, como artigos, livros, relatórios; e à orientação dos integrantes de seu

grupo de pesquisa na coleta e imediata documentação dos dados.

Ao comparar profissionais da informação e pesquisadores, White (2014) observa que ambos fazem uso de diretrizes, sejam padrões formais estabelecidos ou guias de uso individual ou local, sendo que os primeiros possuem mais experiência na utilização de metadados, e os pesquisadores são mais específicos, menos focados na descrição, e mais preocupados com os *softwares* empregados. Para a autora, fica assinalada a distinção no papel do cientista como depositante e do profissional da informação como curador, um aspecto que interfere na arquitetura dos repositórios. As diferenças são esperadas, pela bagagem profissional de cada um dos grupos. Cabe aos mantenedores dos repositórios mapear antecipadamente as necessidades das comunidades-alvo, e conciliar uma equipe de trabalho para estudo e acompanhamento do(s) esquema(s) de metadados adotado(s).

Rousidis *et al.* (2014), em estudo no repositório Dryad, validaram a suposição dos problemas de qualidade ocorrer, majoritariamente, com os elementos Criador, Data e Tipo. Quando focam no campo Assunto, detectam média de 4,79 palavras-chave por pacote de dados, e várias questões de qualidade, como erros de digitação, entrada de informação irrelevante, extrema diversidade de sinônimos, inconsistência entre plural e singular (Rousidis *et al.*, 2015). Os autores sugerem um maior controle manual da entrada de dados, uma das etapas da curadoria.

Em outro estudo, com *datasets* de imagens oftalmológicas, Khan *et al.* (2021) analisam 94 conjuntos de dados em acesso aberto (507.724 imagens, 125 vídeos) e verificam que detalhes técnicos das imagens e sua aquisição foram registrados, porém não ocorreu o mesmo com informação clínica. As características dos pacientes (sexo, idade, origem étnica) constavam em menos de 20% dos *datasets*, o período de coleta em 19% e os critérios de inclusão e exclusão em apenas 15%. Nota-se descuido do depositante com elementos descritivos indubitavelmente reconhecidos por ele, no papel de pesquisador, como essenciais no desenho de qualquer pesquisa, coleta de dados e relato dos resultados. O esperado é que os cientistas mantenham o mesmo zelo ao compartilhar os dados no repositório e ao realizar a pesquisa.

Certamente os padrões melhoram o fluxo de informações dentro das comunidades científicas, mas também podem criar barreiras e impedir a inovação quando são prematuros ou inadequados (Borgman, 2015). Mais uma vez, os mantenedores de repositórios e profissionais da informação têm papel fundamental em avaliar o nível de desenvolvimento do esquema a ser empregado, a necessidade de inclusão de elementos novos que deem conta de uma especificidade institucional e a junção de esquemas especializados a um genérico. A execução de testes-piloto com os cientistas poderá evitar incômodos futuros e retrabalho.

4 OS METADADOS E O PROFISSIONAL DA INFORMAÇÃO

O profissional da informação tem expandido o escopo das suas atividades para se adequar às mudanças tecnológicas e às demandas dos usuários por novos serviços, antes mesmo do surgimento das bibliotecas digitais, logo em seguida com os repositórios institucionais e, mais recentemente, os repositórios de dados de pesquisa, gerando novos significados, revisando conceitos antigos e vendo seu espaço profissional crescer, absorvendo novas funções e desenvolvendo competências.

Esse profissional está capacitado para atuar em ambientes onde o ciclo da informação acontece, desde a seleção e aquisição, passando pela representação descritiva e temática, pelo armazenamento, até a descoberta e disseminação da informação. “Não é o diploma que concede o papel a esse profissional, mas é o mercado o regulador.” (Boeres, 2017, p. 36). Para a mesma autora, o profissional da informação é fruto da sua formação, que congrega conhecimentos da Ciência da Informação, uma Ciência social aplicada, que tem como objetivo estudar todos os aspectos da informação e analisar os processos da sua construção, da sua comunicação e do seu uso. Portanto, o profissional da informação necessita buscar capacitação e atualização constantemente para acompanhar os novos tempos e suas demandas. Foi no final da década de 1990 que surge o chamado Bibliotecário de metadados, refletindo a mudança do papel do Bibliotecário de catalogação.

Em estudo realizado em 2010, Han e Hswe buscaram determinar as competências dos bibliotecários de metadados em comparação com as dos bibliotecários de catalogação. Os dados coletados mostram que o conjunto de habilidades necessárias para um bibliotecário de metadados estão relacionadas com: o conhecimento de padrões de metadados (incluindo MARC e controle de autoridade), o desempenho no trabalho (como analítico, comunicação, interpessoal, organizacional e solução de problemas), e a capacidade de trabalhar em um ambiente de equipe e de forma independente. Em termos de tecnologia da informação, eles devem entender *Extensible Markup Language* (XML), *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) e *Resource Description Framework* (RDF), que são usados para conversão e compartilhamento de metadados. Ainda, os pesquisadores relatam que, a partir de 2004, os bibliotecários de metadados começaram a fazer parte de outros tipos de unidades ou departamentos dentro de uma biblioteca, como uma unidade de biblioteca digital, computação de biblioteca e serviços de mídia, e aquisição e gerenciamento de informações. A inserção de bibliotecários de metadados nesses espaços sugere que a posição de tal bibliotecário se originou em resposta às mudanças na forma como as bibliotecas estavam entregando e fornecendo acesso às suas coleções.

O contexto apresentado, que mostra a evolução do bibliotecário de catalogação em relação ao trabalho com metadados, exige do profissional familiaridade e experiência com uma variedade de formatos, padrões, esquemas,

ferramentas e práticas recomendadas. A posição do bibliotecário de metadados reflete a natureza em rápida mudança da catalogação na Biblioteconomia. Em dois artigos Gold (2007a e 2007b), já apontava como papel dos bibliotecários: criar metadados para descoberta e descrição de dados, organizando documentação para curadoria digital e oferecendo suporte de preservação de dados.

Kennan (2016), em estudo com bibliotecários atuando em universidades nos Estados Unidos, Reino Unido, Canadá e Austrália, listou os requisitos e habilidades indispensáveis para os bibliotecários de dados. O primeiro conjunto inclui habilidades interpessoais e características comportamentais, comunicação de alto nível (incluindo redação de documentação e outras redações técnicas). Outros atributos vinculados ao perfil profissional do bibliotecário de dados são a capacidade de estar continuamente atualizando-se, ter uma "curiosidade sem limites" e ser adaptável. O segundo conjunto de habilidades está relacionado ao conhecimento sobre o ambiente de pesquisa da universidade ou instituição de pesquisa, as políticas das agências financiadoras, as métricas da pesquisa e a avaliação. O terceiro conjunto refere-se a conhecimentos específicos sobre o uso de dados e o entendimento dos tipos de dados (quantitativo, qualitativo), licenças (*Copyright* e *Creative Commons*), preservação de dado e padrões e esquemas de metadados (*Dublin Core*, *Darwin Core*, *RDF*, *DIF*, *ANZLIC*, *RIF-CS*). A maioria dos participantes do estudo de Kennan (2016) enfatizou a necessidade desse conhecimento sobre a finalidade e o uso de metadados e

demonstrou uma vontade de aceitar e aprender os diferentes padrões e esquemas aplicáveis em diferentes contextos.

A gestão de dados de pesquisa abrange uma ampla gama de atividades durante todo o ciclo de vida dos dados. Geralmente, requer um alto nível de interação com os pesquisadores e, também, oferecer serviços de suporte, inclusive técnicos. O bibliotecário de metadados pode trabalhar na alfabetização em dados (padrões de metadados, formatos, etc.), auxiliar a identificar padrões e formatos de dados e padrões de metadados, gerenciar a coleta de dados e a descrição, estudar as melhores práticas para estruturas de dados (tipos, formatos, vocabulários, ontologias e metadados), auxiliar a encontrar informações sobre estruturas de dados, tipos, formatos, vocabulários, ontologias e metadados (Schmidt; Shearer, 2016).

Existem vários modelos de ciclo de vida de dados, como por exemplo o desenvolvido pelo Arquivo de Dados do Reino Unido, o *UK Data Archive Data Lifecycle*, no qual é dada ênfase ao armazenamento de dados bem organizados, bem documentados, preservados e compartilhados para a promoção da investigação científica (UK DATA ARCHIVE, 2012) ou o modelo do Data *Documentation Initiative* (DDI) que desenvolveu o *DDI Combined Lifecycle Model*, um modelo de ciclo de vida para dados de pesquisa, particularmente da área das Ciências Sociais ou ainda, o modelo do *Digital Curation Centre* (DCC), o *Curation Lifecycle Model* é um modelo mais complexo dentre os outros modelos de curadoria digital, preocupando-se em promover uma curadoria e preservação de dados bem-sucedida, a qual tem início desde a

conceituação dos dados até o descarte ou a seleção dos mesmos para reutilização e preservação a longo prazo.

Porém, acreditamos que o modelo de ciclo de vida dos dados elaborado pela *Data Observation Network of Earth* (DataONE), por ter foco mais específico em uma área, os dados da terra, seus oito componentes contemplam mais ampla e detalhadamente as questões relativas à importância dos metadados. A Figura 2 mostra os componentes do *DataONE Data Life Cycle*: planejar, coletar, assegurar, descrever, preservar, descobrir, integrar e analisar.

Figura 2: Ciclo de vida dos dados



Fonte: DataONE, [201-?].

De acordo com as informações da DataONE, existem várias versões de um ciclo de vida de dados com diferenças que variam conforme as práticas entre domínios ou comunidades (DATAONE, [201-?]). Porém, entre os componentes podemos identificar a importância da descrição dos metadados:

- a) planejar o arquivamento dos dados e a descrição dos mesmos: devem ser bem organizados, gerenciados e documentados para garantir que os arquivos possam ser localizados e seu conteúdo seja compreensível. Nesta etapa, é crucial identificar os dados sensíveis;
- b) coletar e prever como os dados serão descritos: os campos de dados devem ser separados de forma consistente por delimitadores aceitáveis (por exemplo, vírgulas) e usar esquemas de codificação consistentes, de forma a facilitar sua preservação e posterior acesso;
- c) assegurar a qualidade dos dados e os métodos de controle devem fazer parte do arquivo de informações de dados, preparar e documentar o plano de gestão de dados para garantir a confiabilidade e precisão. O controle de qualidade é altamente dependente da metodologia de coleta de dados, por esse motivo ela deve ser documentada com clareza;
- d) descrever os dados de forma minuciosa, por meio da adoção de padrões e ferramentas de metadados, além de permitir a interoperabilidade por meio dos metadados, possibilita que outros descubram, compreendam e usem esses dados. É necessário definir um dicionário de dados, no início do projeto, para descrever e documentar os dados com eficiência, por meio de metadados, que descrevem o contexto digital do conjunto de dados, o formato para localização geoespacial, formatos de data e

hora, as etapas de processamento, o conteúdo de um arquivo de dados com as informações sobre a equipe de coleta, o responsável pelo conjunto de dados, os patrocinadores da pesquisa. Ainda, para descrever o contexto científico (como foram coletados, quais os instrumentos usados), fornecer informações sobre os parâmetros (unidade de medida, formato do conjunto de dados, definição dos códigos usados), bem como, informar como o conjunto de dados deve ser citado. As taxonômicas completas para cada espécie devem ser incluídas na descrição dos dados;

- e) preservar é uma etapa do trabalho que é realizada em conjunto com os gestores do repositório de dados, os quais fornecem orientações sobre como gerar os metadados e preservar os dados e quais formatos usar. Não se pode esquecer a necessidade do uso de terminologias padrões (como palavras-chave, ontologias). Nesta etapa, também é necessário identificar a sensibilidade dos dados e fazer a classificação de segurança apropriada;
- f) descobrir configura-se no compartilhamento e publicação dos dados usando ferramentas e serviços baseados na *Web* para que as pessoas possam encontrar os dados e entender o seu conteúdo, assim como identificar outros conjuntos de dados e repositórios que possam complementar e agregar valor à determinada pesquisa. Para facilitar a descoberta, o nome dos arquivos deve ser descritivo e refletir o seu conteúdo;

- g) integrar os dados, a fim de que sejam compatíveis e comparáveis e documentados por meio de metadados. Todas as etapas realizadas para criar o conjunto de dados de uma pesquisa devem ser documentadas de forma completa e precisa para permitir que outros avaliem e interpretem seus resultados. Por exemplo, descrever os parâmetros geoespaciais (resolução, projeção, escala) de dados referenciados espacialmente ao integrar conjuntos de dados geoespaciais de fontes múltiplas;
- h) analisar consiste em integrar dados para que sejam compatíveis e comparáveis com outros. A descrição das etapas de processamento é imprescindível para converter uma observação em um produto de dados derivado. Devem ser documentadas de forma completa e precisa para permitir que outros avaliem e interpretem seus resultados.

Finalmente, cumpre-se o objetivo principal da descrição completa, precisa e consistente de metadados do conjunto de dados de uma pesquisa que, ao serem publicados, utilizando ferramentas e serviços como os repositórios de dados, possibilitam a reprodutibilidade, além de serem descobertos e acessados por qualquer pessoa. Organizar e descrever com metadados apropriados, armazenados em local onde possam ser monitorados e disponibilizados para uso futuro e preparados para migração ou transformação, caso o formato de dados esteja danificado ou obsoleto (Almeida, 2019), também se tornou uma das atividades primordiais dos profissionais da informação ou bibliotecários de metadados.

Mas, por outro lado, “[...] exige o desenvolvimento, a coordenação e investimentos em vários setores da biblioteca para criar um sistema onde, certamente, estarão envolvidos a segurança dos dados, a preservação, o acesso e o controle dos metadados.” (Cunha, 2010, p. 193).

5 CONSIDERAÇÕES FINAIS

Como um dos elementos que compõem o ecossistema da Ciência Aberta, os repositórios tornam-se fontes de busca de dados para pesquisas adicionais e cruzamento de dados, com vários níveis de complexidade para todos os atores envolvidos, porque não são estáticos como uma publicação científica. Apresentam desafios relacionados ao versionamento de arquivos e *softwares* para leitura e processamento, capacidade computacional e interação máquina-máquina, custos para armazenamento de *big data* e manutenção dos repositórios, perda de dados, preparo de *datasets* com a presença de dados pessoais, reusabilidade efetiva dos dados, duplicação de *datasets* em repositórios, indexação dos milhares de repositórios por serviços/mecanismos de busca (por exemplo, *Google Dataset Search*, *Data Citation Index*), certificação internacional, entre outros.

Especificamente para o profissional da informação, que desenvolve atividades focadas tanto no produtor de dados (seleção e avaliação de esquemas de metadados, capacitação e apoio técnico, preenchimento qualificado dos metadados) quanto nos usuários secundários (descoberta fácil dos dados, funcionalidades de busca e navegação do repositório, testes de

usabilidade), consideramos que entre os principais desafios estão a sobreposição de esquemas de metadados e o atendimento de necessidades específicas, de acordo com a área do conhecimento, especialmente em repositórios genéricos, sem perder de vista os princípios FAIR.

Na avaliação dos metadados, cuja atividade entendemos ser da competência do profissional da informação, deve-se considerar a quantidade e qualidade dos mesmos e sua adequação para descoberta de dados. Porém, Löffler *et al.* (2021) salientam que solicitar aos pesquisadores que forneçam “metadados ricos” para descrever seus conjuntos de dados, sem qualquer orientação, não resultará em algo satisfatório, pois muitas vezes não é claro quais são as informações relevantes, para a descoberta dos dados, que devem incluir no momento de submissão dos mesmos no repositório. Os autores recomendam documentar as práticas, principalmente aquelas aplicadas a domínios específicos, fornecer mais informações sobre os vários padrões de metadados e as categorias de pesquisa por campo de busca.

Segundo Almeida (2019), às bibliotecas cabem agora, dentre outros serviços, a definição de padrões de metadados e a realização de treinamentos voltados para a gestão de dados oriundos de pesquisas. Dessa forma, é importante que o profissional da informação esteja próximo aos produtores de dados e engajado ativamente nas atividades de curadoria dos dados e de capacitação e suporte aos pesquisadores, com a finalidade de compartilhar dados de pesquisa em acesso aberto, atendendo aos princípios FAIR, e de modo a qualificar os metadados e, conseqüentemente, os próprios dados. O uso

e reuso dos dados de pesquisa, de forma adequada, estão diretamente relacionados com a qualidade com que são descritos.

REFERÊNCIAS

ALMEIDA, F. G. *Suporte à gestão de dados de pesquisa: uma ampliação dos serviços oferecidos pelas bibliotecas*. 2019. Tese (Doutorado em Gestão & Organização do Conhecimento) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <http://hdl.handle.net/1843/31596>. Acesso em: 19 jul. 2021.

ÁVILA BARRIENTOS, E. Análisis de los principios para la descripción de datos de investigación mediante datacite metadata schema. *Anales de documentación*, Múrcia, v. 23, n. 1, 2020. Disponível em: <http://dx.doi.org/10.6018/analesdoc.400341>. Acesso em: 29 jun. 2021.

BOERES, S. A. de A. *Competências necessárias para equipes de profissionais de preservação digital*. 2017. Tese (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2017. Disponível em: <https://repositorio.unb.br/handle/10482/24354>. Acesso em: 19 jul. 2021.

BORGMAN, C. L. *Big data, little data, no data: scholarship in the networked*. Cambridge: MIT Press, 2015.

CAMPBELL, G. D. Metadata, Metaphor, and Metonymy. In: SMIRAGLIA, Richard. *Metadata: A Cataloger's Primer*. New York: Haworth Information, 2006. p. 4-5.

CHAPMAN, A. *et al.* Dataset search: a survey. *The VLDB Journal*, Heidelberg, v. 29, p. 251-272, 2020. Disponível em: <https://arxiv.org/abs/1901.00735>. Acesso em: 19 jul. 2021.

CORRÊA, F. C. *Gestión de datos de investigación*. Barcelona: Editorial UOC, 2016.

CUNHA, M. B. da. A biblioteca universitária na encruzilhada. *DataGramZero*: revista de ciência da informação, Rio de Janeiro, v. 11, n. 6, dez. 2010. Disponível em: <https://repositorio.unb.br/handle/10482/14869>. Acesso em: 19 jul. 2021.

DATACITE METADATA WORKING GROUP. *DataCite metadata schema documentation for the publication and citation of research data and other research outputs*. Version 4.4. DataCite e.V. 2021. Disponível em: <https://doi.org/10.14454/3w3z-sa82>. Acesso em: 15 jun. 2021.

DATAONE. *Data life cycle*. [201-?]. Disponível em: <https://old.dataone.org/data-life-cycle>. Acesso em: 19 jul. 2021.

GOLD, A. Cyberinfrastructure, Data, and Libraries, Part 1; A Cyberinfrastructure Primer for Librarians. *D-Lib Magazine*, Reston, v. 13 n. 9/10, 2007a. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>. Acesso em: 19 jul. 2021.

GOLD, A. Cyberinfrastructure, data, and libraries, part 2: Libraries and the data challenge: Roles and actions for libraries. *D-Lib Magazine*, Reston, v. 13 n. 9/10, 2007b. Disponível em: <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>. Acesso em: 19 jul. 2021.

HAN, M-J.; HSWE, P. The evolving role of the metadata librarian: competencies found in job descriptions. *Library resources & technical services*, Chicago, v. 54, n. 3, p. 129-141, July 2010. Disponível em: <https://journals.ala.org/index.php/lrts/article/view/5557/6839>. Acesso em: 19 jul. 2021.

KENNAN, M. A. Data management: knowledge and skills required in research, scientific and technical organisations. *In: IFLA WORLD LIBRARY AND INFORMATION CONGRESS, 2016. Columbus. Proceedings...* [S. l.]: IFLA, 2016. Disponível em: <http://library.ifla.org/1466/1/221-kennan-en.pdf>. Acesso em: 19 jul. 2021.

KHAN, S. M. *et al.* A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet digital health*, Oxford, v. 3, n. 1, e51–66, jan., 2021. Disponível em: [https://doi.org/10.1016/S2589-7500\(20\)30240-5](https://doi.org/10.1016/S2589-7500(20)30240-5). Acesso em: 05 jul. 2021.

KOESTEN, L. M. *et al.* The trials and tribulations of working with structured data - a study on information seeking behaviour. *In: CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2017, Denver. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 2017. p. 1277-1289. Disponível em: <https://doi.org/10.1145/3025453.3025838>. Acesso em: 03 jul. 2021.

LÖFFLER, F. *et al.* Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? *PLoS ONE*, San Francisco, v. 16, n. 3, e0246099, 2021. Disponível em: <https://doi.org/10.1371/journal.pone.0246099>. Acesso em: 03 jul. 2021.

MONTEIRO, E. C. de S. de A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados em repositórios de dados de universidades. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Florianópolis, v. 23, n. 53, p. 160-173, set./dez., 2018. Disponível em: <https://doi.org/10.5007/1518-2924.2018v23n53p160>. Acesso em: 28 jun. 2021

NATIONAL INFORMATION STANDARDS ORGANIZATION. *A Framework of guidance for building good digital collections*. 3rd. ed. Baltimore: NISO, 2007. Disponível em: <http://www.niso.org/publications/rp/framework3.pdf>. Acesso em: 05 abr. 2021.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. *OECD principles and guidelines for access to research data from public funding*. Paris: OECD, 2007. Disponível em: <https://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 24 jun. 2021.

PAVÃO, C. M. G. *et al.* Metadados e repositórios institucionais: uma relação indissociável para a qualidade da recuperação e visibilidade da informação. *PontodeAcesso*, Salvador, v. 9, n. 2, p. 103-116, 2015. Disponível em: <http://hdl.handle.net/10183/175147>. Acesso em: 25 jun. 2021.

PONTIKA, N. *et al.* Fostering open science to research using a taxonomy and an eLearning portal. *In: INTERNATIONAL CONFERENCE ON KNOWLEDGE TECHNOLOGIES AND DATA-DRIVEN BUSINESS*, 15., 2015, Graz. *Proceedings [...]*. New York: ACM, 2015. Article n. 11. Disponível em: <https://doi.org/10.1145/2809563.2809571>. Acesso em: 28 jun. 2021.

ROCHA, R. P. da *et al.* *Acesso aberto a dados de pesquisa no Brasil: planejamento para implantação de comunidade*

produtora de dados para o repositório Rede de Dados de Pesquisa. 2019. Disponível em: <http://hdl.handle.net/20.500.11959/1204>. Acesso em: 25 jun. 2021.

ROUSIDIS, D. *et al.* Metadata for big data: a preliminary investigation of metadata quality issues in research data repositories. *Information services and use*, Amsterdam, v. 34, n. 3-4, p. 279-286, 2014. Disponível em: <https://content.iospress.com/download/information-services-and-use/isu746?id=information-services-and-use%2Fisu746>. Acesso em: 19 jun. 2021.

ROUSIDIS, D. *et al.* Evaluation of metadata in research data repositories: the case of the DC.Subject element. *In*: GAROUFALLOU, E.; HARTLEY, R.; GAITANOU, P. (eds). *Metadata and semantics research*. [S. l.]: Springer, 2015. Disponível em: https://doi.org/10.1007/978-3-319-24129-6_18. Acesso em: 19 jun. 2021.

SALES, L. F.; SAYÃO, L. F. *Dados de pesquisa: quem ama cuida*. Brasília: Comissão Nacional de Energia Nuclear: IBICT, 2019. Disponível em: <http://livroaberto.ibict.br/handle/123456789/1083>. Acesso em: 25 jun. 2021.

SCHMIDT, B.; SHEARER, K. *Librarians' Competencies Profile for Research Data Management*. Joint Task Force on Librarians' Competencies in Support of EResearch and Scholarly Communication. 2016. Disponível em: https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf. Acesso em: 19 jul.2021.

SOLODOVNIK, I. Development of a metadata schema describing institutional repository content objects enhanced by "LODE-BD" strategies. *JLIS.it.*, Macerata, v. 4, n. 2, p. 109-

144, 2013. Disponível em: <http://dx.doi.org/10.4403/jlis.it-8792>. Acesso em: 28 jun. 2021.

UNESCO. *Hacia una recomendación de la UNESCO sobre la ciencia abierta: crear un consenso mundial sobre la ciencia abierta*. Paris: UNESCO, 2020. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000373209_spa. Acesso em: 24 jun. 2021.

WHITE, H. C. Descriptive metadata for scientific data repositories: a comparison of information scientist and scientist organizing behaviors. *Journal of library metadata*, New York, v. 14, n. 1, p. 24-51, 2014. Disponível em: <https://doi.org/10.1080/19386389.2014.891896>. Acesso em: 19 jun. 2021.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, Londres, v. 3, article 160018, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 02 jul. 2021.

WU, M. *et al.* Data discovery paradigms: user requirements and recommendations for data repositories. *Data science journal*, London, v. 18, article 3, 2019. Disponível em: <https://doi.org/10.5334/dsj-2019-003>. Acesso em: 19 jun. 2021.