

INFRAESTRUTURA DE CONHECIMENTO E O PAPEL DA BIBLIOTECA DIGITAL NO GERENCIAMENTO DE DADOS

KNOWLEDGE INFRASTRUCTURE AND THE ROLE OF THE DIGITAL LIBRARY IN THE DATA MANAGEMENT

Elizabeth Cristina de Souza de Aguiar Monteiro¹
Ricardo César Gonçalves Sant'Ana²

RESUMO

Introdução: Os dados primários de pesquisas são gerados e coletados por diversos meios e constituem-se como fator chave no seu desenvolvimento, principalmente quando de cunho empírico. Da gestão destes dados emerge a necessidade crescente de planejamento e investimentos em uma infraestrutura de conhecimento específica que pode ter nas bibliotecas digitais um importante componente destes novos cenários. **Objetivo:** Neste artigo são apresentados os principais requisitos à infraestrutura de conhecimento para gestão de dados e o possível papel da biblioteca digital como elemento de apoio. **Metodologia:** Foi utilizada a abordagem qualitativa com revisão de literatura sobre infraestrutura do conhecimento e bibliotecas digitais para a discussão dos temas. **Resultados:** Apresenta-se requisitos para a infraestrutura de conhecimento para gestão de dados e as potenciais contribuições das bibliotecas digitais. **Conclusão:** Os requisitos para infraestrutura de conhecimento apresentam aderência em relação a características das bibliotecas digitais o que pode justificar sua participação na gestão de dados.

Descritores: Infraestrutura do conhecimento. Bibliotecas digitais. Gerenciamento de dados.

¹ Doutoranda em Ciência da Informação pela Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Campus de Marília. Bibliotecária da Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Campus de Marília. E-mail: ecsamonteiro@gmail.com

² Doutor em Ciência da Informação pela Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Campus de Marília. Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Campus de Marília. E-mail: ricardo.santana@unesp.br

1 INTRODUÇÃO

A evolução de instrumentos tecnológicos, o uso de computadores, a internet e os dados digitais³ vêm, ao longo dos anos, transformando a maneira como a pesquisa científica é conduzida e fortalece o protagonismo e o alcance dos dados científicos nesse contexto, criando desafios para a comunidade de pesquisa no gerenciamento e curadoria dos dados (TENOPIR; BIRCH; ALLARD, 2012; SHAKERI; GRACY, 2014).

As tecnologias, das mais variadas, são componentes importantes na forma de se fazer ciência porém, trazem desafios como a necessidade de melhorias nos processos de captura, análise, modelagem, visualização e preservação de dados e informações científicas (GRAY, 2007; MAYER-SCHONBEERGER; CUKIER, 2013).

Como resultado desse avanço, muitas disciplinas científicas vêm ganhando características que colocam em foco o uso intensivo de dados (TENOPIR; BIRCH; ALLARD, 2012; SHAKERI; GRACY, 2014). Por outro lado, existem algumas disciplinas que produzem ou tratam quantidades menores de dados e fazem parte da pequena ciência, também conhecida como cauda longa da ciência (WALLIS; ROLANDO; BORGMAN, 2013).

O cenário atual, e em especial o protagonismo dos dados, contribuem para que se fortaleça a tendência de que as pesquisas se desenvolvam de forma colaborativa, com os pesquisadores participando de infraestruturas de conhecimento que gerenciam o fluxo de trabalho e os conjuntos de dados envolvidos (BORGMAN, et al., 2014).

A infraestrutura de conhecimento pode ser definida como “[...] redes robustas que incluem recursos humanos, artefatos e instituições que geram, compartilham e mantêm o conhecimento específico sobre o ser humano e mundos naturais” (EDWARD, 2012, tradução nossa). Uma infraestrutura de conhecimento pode auxiliar os pesquisadores a identificar pesquisas e

³ A coleta de dados é comumente realizada empregando alguma forma de tecnologia digital, e os dados coletados ou gerados manualmente, quase sempre, são convertidos em um formulário digital para análise (GREENBERG, et al., 2009).

publicações relacionadas a sua temática retroalimentando positivamente o desenvolvimento científico e ampliando o alcance de seus resultados de forma mais ágil e dinâmica (STOCK et al., 2012).

As bibliotecas digitais podem representar um componente com forte potencial de contribuição no gerenciamento de dados quando implantadas no início do ciclo de vida dos projetos de pesquisas (BORGMAN, et al., 2007). Os diversos requisitos funcionais das bibliotecas digitais podem ajudar no gerenciamento dos dados científicos e no apoio técnico aos cientistas contribuindo para a infraestrutura do conhecimento. Conforme o volume e a variedade de dados científicos, surgem novos desafios para a ciência e para a gestão destes dados.

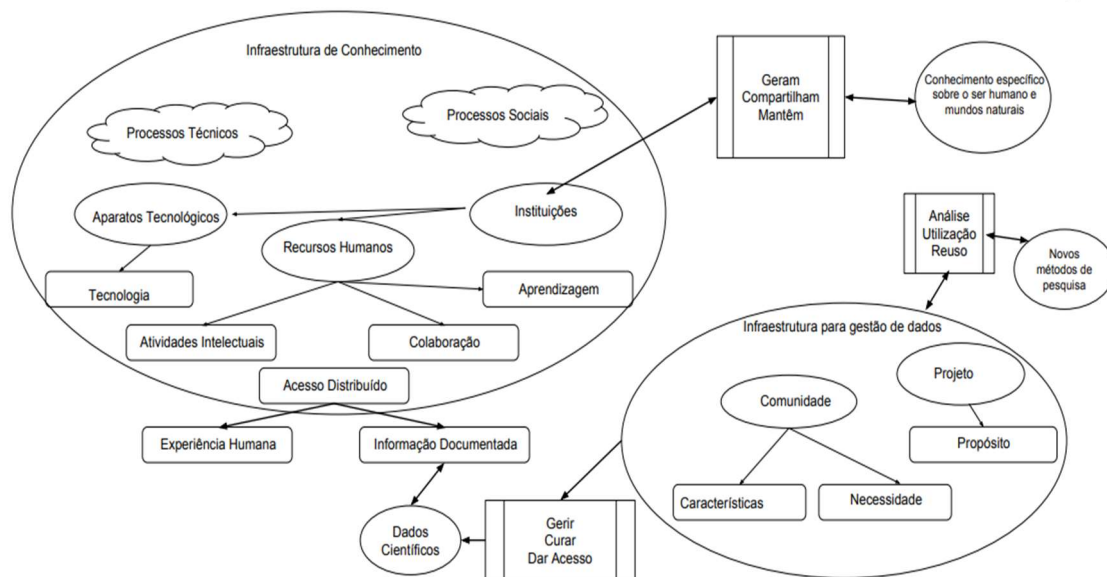
Dessa forma, o objetivo deste artigo é apresentar requisitos à infraestrutura de conhecimento para gestão de dados e o possível papel da biblioteca digital como elemento de apoio.

Como metodologia foi utilizada a abordagem qualitativa com a revisão de literatura sobre infraestrutura do conhecimento, bibliotecas digitais, *small science*, *big science* e *e-science* para a discussão sobre potenciais contribuições da biblioteca digital neste cenário.

2 INFRAESTRUTURA DE CONHECIMENTO

Conforme sintetizado na figura 1, infraestruturas são ecologias ou sistemas adaptativos complexos que “[...] consistem de muitas partes que interagem através de processos sociais e técnicos, com variados graus de sucesso.” e incluem aparatos tecnológicos recursos humanos, colaboração, acesso distribuído e informação documentada (BORGMAN et al., 2015, p. 207, tradução nossa).

Figura 1 - Requisitos de infraestrutura para gerenciamento de dados



Fonte: Elaborado pelos autores

Infraestrutura de conhecimento inclui tecnologia, atividades intelectuais, aprendizagem, colaboração e acesso distribuído à experiência humana e a informação documentada (EDWARDS, 2013). Bibliotecas digitais são elementos importantes na infraestrutura do conhecimento (BORGMAN et al., 2015).

Stock et al. (2012) definiram infraestrutura de conhecimento na área geoespacial que pode ser adequada para as demais áreas do conhecimento, complementando a definição de Edwards (2013) e Stock et al. (2012). Segundo os autores, infraestrutura de conhecimento é um conjunto de componentes interoperáveis, que inclui *software*, *hardware*, informações, procedimentos e padrões que trabalham juntos para apoiar a descoberta científica e criação de recursos incluindo publicações, conjuntos de dados e serviços web (STOCK et al., 2012).

A preparação e implementação de uma infraestrutura para gerenciamento de dados dependem do propósito dos projetos e das características e necessidades de uma comunidade para gerir, curar e dar acesso a seus dados científicos, fazer a análise, utilização e reuso para novos métodos de pesquisas (BORGMAN et al., 2014; BORGMAN et al., 2015).

Para viabilização de uma infraestrutura que visa gerenciar dados são necessários diversos recursos, entre eles os aparatos tecnológicos, os

sistemas de *hardware* e *software*, os recursos humanos com a equipe formada por profissionais multidisciplinares e os pesquisadores, a informação documentada que inclui o Plano de gerenciamento de dados (PGD), os documentos descrevendo os procedimentos na manipulação dos dados e os manuais de serviços que orientam todos os envolvidos, que armazenam, que gerenciam e que reutilizam os dados, a colaboração entre todos os envolvidos no trabalho e na pesquisa e as ferramentas que possibilitam o acesso distribuído. Sua estruturação e sua implementação estão baseadas nos propósitos dos projetos dos pesquisadores e as características e necessidades da comunidade que a infraestrutura vai atender.

Destaca-se que o PGD é o documento elaborado pelos pesquisadores no início do projeto de pesquisa em que são descritos as diretrizes para o ciclo de vida dos dados (MONTEIRO, SANT'ANA, 2018). Esse documento também auxiliará os profissionais que atuam na infraestrutura e na biblioteca digital para gestão e disponibilização dos dados, e os usuários que vão utilizar e replicar os dados.

Nesse cenário, a adoção de uma infraestrutura de conhecimento apoiado a implantação de bibliotecas digitais para o gerenciamento de dados se torna componente importante.

As bibliotecas digitais podem ser implantadas em várias etapas ao longo do ciclo de vida de projetos de pesquisa científica, através da coleta, análise, gerenciamento, documentação, publicação, curadoria e preservação dos dados (BORGMAN et al., 2015).

Santos e Sant'Ana (2013, p. 201) destacam que “O conceito de dado precisa ser redimensionado e se faz necessário que seja entendido e percebido como elemento básico nos fluxos informacionais [...]”

Sendo assim, as bibliotecas digitais de dados e de documentos científicos podem facilitar a colaboração entre pesquisadores, promovendo o progresso da ciência (BORGMAN; WALLIS; ENYEDY, 2007). Métodos científicos e a organização do trabalho colaborativo estão adaptando-se aos volumes e diversidade de dados a serem gerados. Como os dados são combinados a partir de várias fontes e podem resultar em novas interpretações,

o desafio da concepção de estratégias eficazes de gestão de dados se multiplica (BORGMAN et al., 2014).

Dessa forma, a construção de estruturas de conhecimento com apoio das bibliotecas digitais para o gerenciamento desses dados contribui para uma melhor recuperação, acesso, reutilização e preservação dos dados.

3 AS BIBLIOTECAS DIGITAIS

As universidades, centros de pesquisas e laboratórios são ambientes de coleta, produção, processamento, análise, transferência, uso e reuso de dados e informações, tendem a incentivar e apoiar as comunidades científicas no desenvolvimento, divulgação e publicação dos resultados de suas pesquisas (MONTEIRO; SANT'ANA, SANTARÉM SEGUNDO, 2016).

As atividades científicas, em muitos campos, geram dados que aumenta em volume, velocidade e variedade envolvendo ciclos múltiplos. Com a grande quantidade de dados e com o intuito de racionalizar recursos e promover a reprodutibilidade dos dados, agências de financiamento de países como Estados Unidos e Europa exigem a disponibilização dos dados científicos para financiamento dos projetos (BORGMAN et al., 2015).

O gerenciamento de conjuntos de dados é realidade nos domínios da ciência e nos processos de investigação científica independente da quantidade e complexidade dos dados ou área do conhecimento. As bibliotecas digitais podem contribuir como elemento de apoio para a gestão dados.

Bibliotecas digitais são organizações que fornecem os recursos, inclusive o pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e garantir a permanência no tempo de coleções de obras digitais, de modo que estejam acessíveis, pronta e economicamente, para serem usadas por uma comunidade determinada ou por um conjunto de comunidades (DIGITAL LIBRARY FEDERATION, c2004, tradução nossa).

Geralmente, biblioteca digital é planejada, descrita e formalizada em um documento que tem descrito sua finalidade, os elementos que a compõem e as atividades que serão realizadas e desenvolverá sua coleção de acordo com a

missão estratégica da sua instituição que, por conseguinte, refletirá nas características de sua comunidade (TAMMARO; SALARELLI, 2008).

As bibliotecas digitais adaptam-se às necessidades das comunidades as quais oferecem serviços e assumem “[...] a função de um laboratório onde as diversas ‘partes interessadas’ são estimuladas a colaborar para atingir o objetivo comum.” (TAMMARO; SALARELLI, 2008, p. 136). Para desenvolver as atividades, as bibliotecas digitais são compostas por vários requisitos técnicos.

3.1 Requisitos técnicos da biblioteca digital

Segundo Borgman et al. (2005) para que as bibliotecas digitais possam gerenciar dados científicos, os requisitos variam consideravelmente pela escala do projeto científico ao longo de muitas dimensões que podem incluir: ciclo de vida dos dados, volume dos dados, domínio de investigação e os tipos e graus de compreensão.

As bibliotecas digitais podem ser implantadas durante todo o ciclo de vida dos dados para o seu gerenciamento (BORGMAN et al., 2015). Borgman, Wallis e Enyedy (2007) apontam os principais requisitos para sua implantação:

- especificação de uma estrutura padrão de comunicação (como é fornecido pelo XML) para o comunicação e troca de metadados, tanto entre a membros da comunidade diretamente ligada à pesquisa, quanto entre a comunidade heterogêneas. O XML é uma linguagem de descrição para estruturação de documentos;
- especificação da semântica (significado) e sintaxe (estrutura) de um padrão de metadados para utilização por todos os membros das comunidades;
- a implementação de ferramentas que permitam aos membros de múltiplas comunidades supervisionar a criação (manual, semi-automática ou automática) de metadados, bem como a análise, utilização e preservação dos dados.

Stock et al. (2012) destacam que a semântica pode ser usada para tornar a pesquisa sobre os recursos científicos mais inteligentes e apoiar novas formas de inferir e visualizar conhecimentos científicos. A semântica

acrescenta definições de vocabulário, conceitos e termos claros e codificados para a máquina e explica as inter-relações entre eles em formas declaratórias (afirmações) e condicionais (lógicas ou baseadas em regras) (FOX; HENDLER, 2011).

A semântica pode ser alcançada com o uso de ontologias. “À medida que crescem o volume, a complexidade e a heterogeneidade dos recursos de dados, os cientistas precisam cada vez mais de novas capacidades, que dependem de novas abordagens ‘semânticas’ (por exemplo, sob a forma de **ontologias** [...])” (FOX; HENDLER, 2011, p. 159, grifo do autor).

De acordo com o consórcio W3C (2004), ontologia é a definição dos termos utilizados para descrição e representação de uma área do conhecimento.

As ontologias definem um vocabulário comum em uma área específica de domínio fornecendo o conhecimento estruturado e uma infraestrutura para integrar bases de conhecimento. Dessa forma a estrutura ontológica pode compartilhar um entendimento comum da estrutura de informação entre as pessoas ou agentes de *software*, permitir a reutilização de conhecimento de domínio, estabelecer as premissas de domínios explícitos, separar o conhecimento do domínio do conhecimento operacional e analisar o conhecimento de domínio (NOY; McGUINNESS, [200-]).

O uso de metadados é outro requisito apontado para as bibliotecas digitais. Metadados são dados ou informações descritivas ou contextuais que se referem ou estão associados a outro recurso ou objeto (DIGITAL CURATION CENTRE (DCC), c2004-2015a).

Esses metadados, formando um conjunto estruturado de elementos chamado de padrão de metadados, descrevem o recurso e auxilia em sua identificação, localização e recuperação pelos usuários (DCC, c2004-2015a).

Assim como o uso de metadados para descrever os dados nas bibliotecas digitais, o Plano de gerenciamento dos dados, documento que descreve a proveniência dos dados e os passos seguidos durante a coleta, análise e processamento dos dados, também é necessário tanto para os profissionais que atuam na biblioteca que o utilizará para a descrição e

gerenciamento dos conjuntos dos dados quanto para os pesquisadores que vão utilizar e replicar os dados.

Considerando a grande quantidade de dados, sua heterogeneidade e a complexidade em seu tratamento, cada etapa na manipulação dos dados requer o conhecimento dos passos que foram dados anteriormente pois os detalhes documentados auxiliam na interpretação, na manipulação e reutilização dos dados por outros pesquisadores (BORGMAN; DARCH; SANDS, 2015). A reutilização pode ser ampliada com compartilhamento dos dados armazenados.

Outro requisito é a interoperabilidade entre os sistemas. A interoperabilidade é viabilizada através de infraestrutura tecnológica, utilização de padrões abertos e com ocorrência de relacionamentos semânticos assegurando a viabilidade de intercâmbio de metadados pois permite a troca de dados e informações entre sistemas, bases de dados, repositórios e bibliotecas digitais e, assim, facilita a difusão eficiente entre os conteúdos. Essa interoperabilidade pode ser atingida pelo uso do *Digital Object Identifier* (DOI) (SAYÃO, 2007).

A possibilidade de acesso distribuído dos dados científicos é outro ponto importante das bibliotecas digitais. Para esse acesso ocorrer, são necessários os padrões *Reference Model for Open Archival Information Systems* (OAIS), *Open Archives Initiative Protocols for Metadata Harvesting* (OAI-PMH), OpenURL, the Info URI scheme e *Object Reuse and Exchange* (ORE) (BORGMAN, WALLIS, ENYEDY, 2007).

O modelo de referência *Open Archives Information System* (OAIS) é uma iniciativa da *International Organization for Standardization* (ISO), uma norma ISO 14721:2003 desde junho de 2003. Foi publicada pelo *Consultive Committee for Space Data Systems* (CCSDS). Define “[...] um alto nível de modelo de referência para arquivos que precisam de uma preservação de longo prazo” (ARELLANO, 2004, p. 19).

O OAIS é um esquema conceitual “[...] que disciplina e orienta um sistema para a preservação e manutenção do acesso à informação digital por longo prazo” (THOMAZ; SOARES, 2004, p. 17).

Segundo Thomas e Soares (2004) os objetivos do OAIS são:

- ampliar a consciência e a compreensão dos conceitos relevantes para a preservação de objetos digitais, especialmente entre instituições não arquivísticas;
- definir terminologias e conceitos para descrever e comparar modelos de dados e arquiteturas de arquivos;
- ampliar o consenso sobre os elementos e os processos relacionados à preservação e acesso à informação digital;
- criar um esquema para orientar a identificação e o desenvolvimento de padrões.

Outro requisito para o acesso distribuído é o *Open Archive Initiative – Protocol for Metadata Harvesting* (OAI-PMH) - Protocolo da Iniciativa dos Arquivos Abertos. OAI-PMH é um protocolo para coleta automática de metadados entre bibliotecas e repositórios digitais.

O protocolo OAI-PMH

[...] é um mecanismo para transferência de dados entre bibliotecas digitais. É uma interface que um servidor de rede pode empregar para que os metadados de objetos residentes no servidor estejam disponíveis para aplicações externas que desejam coletar esses dados (PROTOCOLO OAI-PMH... [2006]).

O OAI-PMH suporta múltiplos formatos para representar os metadados. Inicialmente o padrão de metadados obrigatório era o Dublin Core não qualificado evoluindo para outros padrões como o Marc 21 (VAN DE SOMPEL; YOUNG; HICKEY, 2003).

Outro fator para que ocorra o acesso distribuído aos dados científicos é o *OpenURL*. *OpenURL* é uma infraestrutura definida pela norma da *National Information Standards Organization* (NISO) que auxilia os usuários da internet a localizar uma cópia do recurso informacional através do link que ele tem permissão de acesso (SAYÃO, 2007). Geralmente essa tarefa é designada à biblioteca que disponibiliza opções dos links para o usuário escolher àquela que tem permissão de acesso de acordo com seu contexto - tecnologia sensível ao contexto, ou seja, tecnologia *Open reference linking* e

[...] define um protocolo para interoperabilidade entre um recurso de informação e um componente de serviço, denominado servidor de links; este, oferece serviços de localização sensíveis ao contexto, através da interpretação de metadados bibliográficos, codificados segundo uma sintaxe própria, que descrevem uma publicação. (SAYÃO, 2007, p. 78).

O terceiro requisito é o uso do *Info URI scheme*. De acordo com Van de Sompel et al. (2006) *Info URI scheme*, facilita a referência dos ativos de informação que têm identificadores nos *namespaces* públicos por meio de URIs. *Info URI scheme* permite *namespaces* públicos que não fazem parte da atribuição URI a ser representado pela dotação e prevê um mecanismo de transição para permitir que *namespaces* públicos se tornarem parte da alocação URI (VAN DE SOMPEL et al., 2006).

O último requisito apontado é *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE) que define normas para a designação e a troca de agregações de recursos da Web. Estas agregações, às vezes chamados de objetos digitais compostos, podem combinar recursos distribuídos com vários tipos de mídia, incluindo textos, imagens, dados e vídeos. O objetivo destas normas é expor o conteúdo nestas agregações para aplicativos que suportam a criação, depósito, troca, visualização, reutilização e preservação (OPEN ARCHIVES... [20--]).

O quadro 2 apresenta uma visão geral dos requisitos técnicos mínimos de uma biblioteca digital.

Quadro 1 – Requisitos técnicos para a biblioteca digital

BIBLIOTECA DIGITAL	REQUISITOS	INSTRUMENTALIZAÇÃO	EXEMPLO
	Software	Gerenciamento da coleção	My Library, Nou-Rau, Aleph, Sophia
	Banco de dados	Armazenamento de dos dados/e gerenciamento de banco de dados	Oracle, PostgreSQ
	Linguagem de descrição dos dados	Linguagem de descrição para estruturação de documentos	XML
	Representação semântica dos dados	Definição de vocabulário ou ontologias	OWL

Representação dos dados	Padrões de metadados	Dublin Core, MARC 21
Interoperabilidade entre sistemas	Viabilidade de interoperabilidade de metadados	OAI-PMH
Recuperação de dados	Viabilidade de recuperação e transferência de dados em formato bibliográfico	Protocolo Z39.50
Intercâmbio de metadados	Viabilidade de intercâmbio de metadados	XML, RDF
Requisito distribuído	Preservação	Open Archival Information System (OAIS)
	Coleta automática de metadados	Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
	Link de acesso	OpenURL
	Referenciação (URI)	INFO URI SCHEME

Fonte: Elaborado pelos autores

O gerenciamento de dados de pesquisa é difícil e, tornar esses dados úteis e compreensíveis para outros pesquisadores que não fizeram parte do processo de coleta e análise ou pesquisadores de domínios diversos que utilizarão os dados para fins não previstos, é muito mais difícil (BORGMAN et al., 2015). Combinar dados de múltiplas fontes para novas interpretações exige uma infraestrutura que pode ser apoiada por bibliotecas, sendo o gerenciamento de dados preponderantes nesse contexto.

4 RESULTADOS E DISCUSSÃO

A importância do gerenciamento de dados científicos tem sido enfatizada nas últimas décadas e tornou-se obrigatório para pesquisadores que solicitam financiamento por agências de fomento nos Estados Unidos e na Europa. No Brasil, o gerenciamento de dados desponta no Estado de São Paulo com a chamada do edital da Fapes, *Chamada de Proposta de Pesquisa - 2017 em seu Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais (PFPMCG)* (FAPESP, 2017).

Trabalho colaborativo sempre esteve presente entre pesquisadores e o compartilhamento de dados científicos tornou-se inevitável nas ciências

atendidas por iniciativas de infraestrutura disciplinar e nacional (CRAGIN et al., 2010).

A infraestrutura de conhecimento apoiada por biblioteca digital contribui para o gerenciamento e o compartilhamento de dados. Conforme visto na figura 1 e no quadro 1, uma infraestrutura para gerenciamento de dados científicos e uma biblioteca digital possuem vários requisitos que estão reunidos no quadro 2.

Quadro 2 - Requisitos de infraestrutura e biblioteca digital

INFRAESTRUTURA	BIBLIOTECA DIGITAL
Aparatos Tecnológicos	Banco de Dados Software
	Linguagem de descrição Representação semântica Interoperabilidade Recuperação Intercâmbio de metadados
Recursos humanos	Pessoal especializado
Informação documentada	PGD, manuais de serviço, projeto
Colaboração	Colaboração
Acesso distribuído	OAIS OAI-PMH OpenURL INFO URI SCHEMA
Projeto - Proposta	Projeto - Escala
Características da comunidade	Missão da instituição
Necessidades da comunidade	Necessidades da comunidade
	Dimensão Ciclo de vida dos dados
	Dimensão Volume dos dados
	Dimensão Domínio da investigação
	Dimensão Tipos e graus de compreensão

Fonte: Elaborado pelos autores.

O quadro 2 apresenta uma visão geral sobre os requisitos da infraestrutura do conhecimento para gestão de dados científicos e da biblioteca

digital. Observa-se um alto nível de aderência entre os requisitos da infraestrutura de conhecimento para gestão de dados e os requisitos da biblioteca digital. Uma análise mais detalhada entre os elementos de cada infraestrutura mostra que a biblioteca digital dispõe dos requisitos necessários para a infraestrutura de conhecimento e de mais quatro requisitos que envolvem:

- os dados (Dimensão Ciclo de vida dos dados e Dimensão Volume dos dados) que forma a coleção da biblioteca;
- e os pesquisadores (Dimensão Domínio da investigação e Dimensão Tipos e graus de compreensão) que são usuários da biblioteca e os criadores do conteúdo que fará parte da coleção.

Esses requisitos correspondem ao que os autores Tamaro e Salarelli (2008, p. 146) descrevem: “Os sistemas de biblioteca digital, sua representação na comunidade pertinente, e o uso que dela faz o usuário se influenciam reciprocamente.” e complementam dizendo que a biblioteca digital é considerada “[...] um produto da sociedade e a importância que é atribuída à organização do conhecimento tem um significado não só econômico, mas um valor pessoal e social.”

Outro aspecto de destaque é que uma infraestrutura de conhecimento evidencia o trabalho colaborativo contribuindo para a reutilização de dados. Esse ponto condiz com a concepção da biblioteca digital que terá em sua coleção conjuntos de dados gerados pela sua comunidade e disponibilizados para utilização por outros pesquisadores.

5 CONSIDERAÇÕES FINAIS

O gerenciamento de dados para o compartilhamento entre equipes e pesquisadores assume diversas proporções e finalidades, incluindo a possibilidade de outros pesquisadores conduzirem diferentes linhas de pesquisas a partir da reutilização dos dados, evitar repetição e retrabalho na coleta e uso de financiamentos, muitas vezes, com dinheiro público. O gerenciamento de dados assume diversas finalidades e proporções e necessita

de planejamento e investimentos em uma infraestrutura do conhecimento específica.

Desta forma, o projeto de infraestrutura de conhecimento com o apoio de bibliotecas digitais terá sucesso com a utilização do potencial de desempenho, consistência e estabilidade de todo o sistema. Serviços de gerenciamento de dados em bibliotecas digitais são oportunos e significativos para auxiliar os pesquisadores.

As bibliotecas digitais servem de apoio desde o início do ciclo de vida dos dados e de informações de cada projeto. A ação inicial no desenvolvimento de ferramentas e serviços para a biblioteca digital para apoiar o ciclo de vida dos dados é identificar os estágios desse ciclo. O ciclo de vida de um determinado projeto pode gerar vários documentos que servem de registro dos métodos utilizados no ciclo e que serão armazenados nas bibliotecas digitais. Esses documentos podem fornecer detalhes suficientes para replicação da pesquisa. Os requisitos para a biblioteca digital podem variar de acordo com o ciclo de vida dos dados da comunidade que atende.

REFERÊNCIAS

ALVARO, E. et al. E-science librarianship: field undefined. **Issues in Science & Technology Librarianship**, Chicago, n. 66, p. 28-43, Summer 2011.

Disponível em: <<http://www.istl.org/11-summer/article1.html>>. Acesso em: 8 set. 2016.

ARELLANO, M. A. Preservação de documentos digitais. **Ciência da Informação**, Brasília, DF, v. 33, n. 2, p.15-27, maio/ago. 2004.

BELL, G. Prefácio. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma**: descobertas científicas na era da eScience. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 11-15.

BORGMAN, et al. Drowning in data: digital library architecture to support scientific use of embedded sensor networks. In: ACM/IEEE Joint Conference on Digital Libraries, JCDL, 2007, Vancouver. **Proceedings...**, [S.l.]: Researchgate, 2007. Disponível em: <<https://www.researchgate.net/publication/220924065>>. Acesso em: 2 nov. 2015.

BORGMAN, C. L. WALLIS, J. C.; ENYEDY, N. Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. **International Journal on Digital Libraries**, New York, v. 7, n. 1, 17-30, Oct. 2007. Disponível em: <<http://link.springer.com/article/10.1007%2Fs00799-007-0022-9#/page-1>>. Acesso em: 12 jan. 2016.

BORGMAN, C. L.; et al. **The ups and downs of knowledge infrastructures in science**: implications for data management. 2014. Disponível em: <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6970177>>. Acesso em: 7 dez. 2015.

BORGMAN, C. L. et al. Knowledge infrastructures in science: data, diversity, and digital libraries. **International Journal on Digital Libraries**, New York, v. 16, n. 3, p. 207–227, Sep. 2015. disponível em: <<http://link.springer.com/article/10.1007%2Fs00799-015-0157-z#/page-1>>. Acesso em: 2 nov. 2015.

CARLSON, J. R.; GARRITANO, J. R. **E-science, cyberinfrastructure and the changing face of scholarship**: organizing for new models of research Support at the Purdue University Libraries. 2010. Disponível em: <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1170&context=lib_research>. Acesso em: 2 set. 2016.

CESAR JUNIOR, R. M. Apresentação à edição brasileira. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma**: descobertas científicas na era da eScience. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 7-8.

CRAGIN, M. H. et al. Data sharing, small science and institutional repositories. **Philosophical Transactions of the Royal Society**, London, n. 368, p. 4023-4038, 2010. Disponível em: <<http://rsta.royalsocietypublishing.org/>>. Acesso em: 17 Jun. 2016.

DIGITAL CURATION CENTRE. **What are metadata standards**. Edinburgh, c2004-2015a. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>>. Acesso em: 31 Jul. 2015.

DIGITAL LIBRARY FEDERATION. **A working definition of digital library [1998]**. Washington, c2004. Disponível em: <<https://old.diglib.org/about/dldefinition.htm>>. Acesso em: 20 fev. 2018.

EDWARDS, P. N. *et al.* **Knowledge infrastructures**: intellectual frameworks and research challenges. Ann Arbor: Deep Blue, 2013. Disponível em: <<http://hdl.handle.net/2027.42/97552>>. Acesso em: 2 nov. 2015.

FOX, P.; HENDLER, J. eScience semântica: o significado codificado na próxima geração de ciência digitalmente apropriada. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma**: descobertas científicas na era da eScience. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 159-163.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais - Chamada de Propostas de Pesquisa 2017**. São Paulo, 2017. Disponível em: <<http://www.fapesp.br/11068>>. Acesso em: 15 dez. 2017.

GRAY, J. **eScience**: a transformed scientific method. 2007. Disponível em: <http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt>. Acesso em: 18 ago. 2015.

GREENBERG, J.; et al. A metadata best practice for a scientific data repository. **Journal of Library Metadata**, Philadelphia, v. 9, n. 3-4, p. 194-212, Dec. 2009. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/19386380903405090>>. Acesso em: 22 nov. 2017.

MAYER- SCHONBERGER, V.; CUKIER, K. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

MEDEIROS, J. C.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 311-322, set. 2012. Disponível em: <<http://www.ibict.br/liinc> >. Acesso em: 8 abr. 2015.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. SANTARÉM SEGUNDO, J. E. e-SCIENCE SEMÂNTICA: integração dos dados na comunicação científica. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 9-29, 2016. Disponível em: <<http://www.periodicos.ufc.br/index.php/informacaoempauta/article/view/2942/688>>. Acesso em: 18 ago. 2016.

MONTEIRO, E. C. S. 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, 2018. No Prelo.

NATIONAL SCIENCE BOARD. **Long-lived digital data collections**: enabling research and education in the 21st century. Alexandria, National Science foundation, 2005. Disponível em: <<http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>>. Acesso em: 30 maio 2016.

NOY, N. F.; McGUINNESS, D. L. **Ontology development 101**: a guide to creating your first ontology. Stanford: Protégé, [200-]. Disponível em: <http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html>. Acesso em: 6 Jul. 2015.

OPEN ARCHIVES *Initiative object reuse and exchange*. [20--]. Disponível em: <<http://www.openarchives.org/ore/>>. Acesso em 15 jan. 2015.

PROTOCOLO OAI-PMH. [2006]. Disponível em: <<http://clube-oai.incubadora.fapesp.br/portal/prot-oai>>. Acesso em: 14 jul. 2009.

SANTOS, P. L. V. A. C.; SANT'ANA, R. C. G. Transferência da informação: análise para valoração de unidades de conhecimento. **DataGramZero**: Revista de Ciência da Informação, Rio de Janeiro, v. 3, n. 2, abr. 2002. Disponível em: <http://www.dgz.org.br/abr02/Art_02.htm>. Acesso em: 01 maio 2015.

SANTOS, P. L. V. A. C.; SANT'ANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela ciência da informação. **Ciência da Informação**, Brasília, DF, v. 42, n. 2, p. 199-209, maio/ago, 2013. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/2286/1996>>. Acesso em: 7 dez. 2015.

SAYÃO, L. F. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL. **Transinformação**, Campinas, v. 19, n. 1, p. 65-82, 2007. <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/621>>.

SHAKERI, S., GRACY, K. F. A model for data curation research in small sciences. **American Society for Information Science and Technology**, Silver Springer, v. 51, n. 1, p. 1-5, 2014. Disponível em: <<http://onlinelibrary-wiley.ez87.periodicos.capes.gov.br/doi/10.1002/meet.2014.14505101094/full>>. Acesso em: 07 fev. 2018.

STOCK, K. et al. To ontologise or not to ontologise: An information model for a geospatial knowledge infrastructure. **Computers & Geosciences**, New York, v. 45, p. 98-108, Aug. 2012. Disponível em: <<http://www-sciencedirect-com.ez87.periodicos.capes.gov.br/science/article/pii/S0098300411003566>>. Acesso em: 05 jan. 2015.

THOMAZ, K.; SOARES, A.J. A preservação digital e o modelo de referência Open Archival Information. **DataGramZero**, Rio de Janeiro, v. 5, n. 1, p. 1-35, fev. 2004.

TOLLE, K.; TANSLEY, S.; HEY, T. Jim Gray e a eScience: um método científico transformado. In: HEY, T.; STEWARD, T.; TOLLE, K. (Org.). **O quarto paradigma: descobertas científicas na era da eScience**. Tradução Leda Beck. São Paulo: Oficina de textos, 2011. p. 17-29.

VAN DE SOMPEL, H.; YOUNG, J. A.; HICKEY, T. B. Using the OAIPMH... differently. **D-Lib Magazine**, [S.l.], v. 9, n. 7/8, Jul./Aug. 2003. Disponível em: <<http://www.dlib.org/dlib/july03/young/07young.html>>. Acesso em: 14 jan. 2015.

VAN DE SOMPEL, H. V. et al. **The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces**. 2006. Disponível em: <<https://www.rfc-editor.org/rfc/pdf/rfc4452.txt.pdf>>. Acesso em: 15 jan. 2015.

WALLIS, J. C.; ROLANDO, E.; BORGMAN, C. L. If we share data, will anyone use them? Data Sharing and Reuse in the Long Tail of Science and Technology. **PLOS ONE**, San Francisco, v. 8, n. 7, p. 1-17, 2013.

W3C. **OWL Web ontology language**. Cambridge, 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em 10 ago. 2015.

KNOWLEDGE INFRASTRUCTURE AND THE ROLE OF THE DIGITAL LIBRARY IN THE DATA MANAGEMENT

ABSTRACT

Introduction: The primary research data are generated and collected by various means and constitute a key factor in its development, especially when empirical. From the management of this data emerges the growing need for planning and investments in a specific knowledge infrastructure that can have in digital libraries an important component of these new scenarios. **Objective:** This article presents the main requirements for the knowledge infrastructure for data management and the possible role of the digital library as a support element. **Methodology:** We used bibliographical research on knowledge infrastructure and digital libraries with a qualitative approach to the discussion of themes. **Results:** Requirements for knowledge infrastructure for data management and the potential contributions of digital libraries are presented. **Conclusion:** The requirements for knowledge infrastructure present adherence to characteristics of digital libraries, which may justify their participation in data management.

Descriptors: Knowledge infrastructure. Digital libraries. Data management.

KNOWLEDGE INFRASTRUCTURE Y EL ROL DE LA BIBLIOTECA DIGITAL EN LA GESTIÓN DE DATOS

RESUMEN

Introducción: Los datos primarios de investigación son generados y recolectados por diversos medios y se constituyen como un factor clave en su desarrollo, principalmente cuando de carácter empírico. De la gestión de estos datos surge la necesidad creciente de planificación e inversiones en una infraestructura de conocimiento

específica que puede tener en las bibliotecas digitales un importante componente de estos nuevos escenarios. **Objetivo:** En este artículo se presentan los principales requisitos a la infraestructura de conocimiento para gestión de datos y el posible papel de la biblioteca digital como elemento de apoyo. **Metodología:** Se utilizó la investigación bibliográfica sobre infraestructura del conocimiento y bibliotecas digitales con abordaje cualitativo para la discusión de los temas. **Resultados:** Se presentan requisitos para la infraestructura de conocimiento para gestión de datos y las posibles contribuciones de las bibliotecas digitales. **Conclusión:** Los requisitos para la infraestructura de conocimiento se adhieren a las características de las bibliotecas digitales, lo que puede justificar su participación en la gestión de datos.

Descriptores: Infraestructura de conocimiento. Bibliotecas electrónicas. *Gestión de datos.*

Recebido em: 24.08.2018

Aceito em: 02.10.2018