



UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

MARIA LÍGIA TRIQUES

A DIMENSÃO RELACIONAL ENTRE CURADORIA DIGITAL E METADADOS

São Carlos/SP
2020

MARIA LÍGIA TRIQUES

A DIMENSÃO RELACIONAL ENTRE CURADORIA DIGITAL E METADADOS

RELATIONAL DIMENSION BETWEEN DIGITAL CURATION AND METADATA

Dissertação apresentada à banca de defesa de Mestrado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos, como requisito para obtenção do título de mestre em Ciência da Informação.

Área de Concentração: Conhecimento, Tecnologia e Inovação

Linha de Pesquisa: Tecnologia, Informação e Representação

Orientador(a): Profa. Dra. Ana Carolina Simionato Arakaki

Agência Financiadora: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)



grupo de pesquisa
dados e metadados

MARIA LÍGIA TRIQUES

A DIMENSÃO RELACIONAL ENTRE CURADORIA DIGITAL E METADADOS

Dissertação apresentada à banca de defesa de Mestrado do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de São Carlos, como requisito para obtenção do título de mestre em Ciência da Informação.

Área de Concentração: Conhecimento, Tecnologia e Inovação

Linha de Pesquisa: Tecnologia, Informação e Representação

Data da defesa: 05 de março de 2020

BANCA EXAMINADORA

Profa. Dra. Ana Carolina Simionato Arakaki (Orientadora)

Docente do Departamento e do Programa de Pós-Graduação em Ciência da Informação da UFSCar

Prof. Dr. Fabiano Ferreira de Castro

Docente do Departamento e do Programa de Pós-Graduação em Ciência da Informação da UFSCar

Profa. Dra. Ana Cristina de Albuquerque

Docente do Departamento e do Programa de Pós-Graduação em Ciência da Informação da UEL

Prof. Dr. Felipe Augusto Arakaki

Docente do Curso de Biblioteconomia da UnB

SUPLENTES:

Prof. Dr. Rogério Aparecido Sá Ramalho

Docente do Departamento e do Programa de Pós-Graduação em Ciência da Informação da UFSCar

Profa. Dra. Paula Regina Ventura Amorim Gonzalez

Docente do Departamento de Ciência da Informação da UFES



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Maria Lígia Triques, realizada em 05/03/2020:

Prof. Dr. Fabiano Ferreira de Castro
UFSCar

Profa. Dra. Ana Carolina Simionato Arakaki
UFSCar

Profa. Dra. Ana Cristina de Albuquerque
UEL

Prof. Dr. Felipe Augusto Arakaki
UnB

AGRADECIMENTOS

Agradeço primeiramente a minha família, que com muito amor e paciência tem me ensinado os caminhos da vida e também a todos que colaboraram direta ou indiretamente na realização deste trabalho.

Á minha mãe Regina que sempre com paciência e carinho tem me mostrado a melhor forma de ver e fazer as coisas, sem nunca desistir.

Ao meu pai Fernando que com muitos conselhos e ensinamentos sempre me ajudou a superar as etapas.

Á minha irmã Carolina que cresceu junto comigo e sempre esteve ao meu lado, nas horas boas e ruins.

Ás minhas adoradas Nika e Lana, que me alegram todos os dias.

Á minha orientadora, Professora Ana Carolina, pelos anos de aprendizado em aulas, orientações, conversas e, principalmente, por ser um exemplo de mulher, professora e pesquisadora. Obrigada pela paciência, amizade e confiança depositadas em mim.

Á Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro nesses dois anos de mestrado para o desenvolvimento dessa pesquisa.

Aos professores do PPGCI pelo compartilhamento de seus conhecimentos.

Aos funcionários do Departamento de Ciência da Informação pela dedicação.

Aos colegas do PPGCI pelos momentos de troca de conhecimento, discussões e também de descontração e risadas.

Ao Grupo de Pesquisa Dados e Metadados que proporcionou grandes ensinamentos e oportunidades.

Aos meus queridos amigos que estiveram comigo e me apoiaram nessa caminhada.

Agradeço também a banca examinadora, composta pelos professores Fabiano, Ana Cristina e Felipe pela participação e leitura, como também pelas sugestões e correções para o trabalho.

Por fim, agradeço à Universidade Federal de São Carlos que proporcionou momentos e oportunidades que levarei por toda a minha vida.

*“Em uma visão extrema, o mundo pode ser visto apenas
como conexões, nada mais.”*

Tim Berners-Lee

RESUMO

O volume e a variedade de dados relativos aos patrimônios culturais têm crescido significativamente nos ambientes digitais, criando diferentes possibilidades e desafios relacionados à representação da informação, sobretudo aos estudos voltados para a curadoria digital. No ambiente *Web*, verifica-se a necessidade de que o tratamento leve em consideração, além de todos os estágios pelos quais os dados circulam, a capacidade de garantir seu contexto informacional. Nesse cenário, a curadoria digital ganha destaque como uma atividade interdisciplinar que envolve não só a manutenção e a preservação dos recursos informacionais, mas também a capacidade desses recursos estarem continuamente coerentes com os seus propósitos informacionais. Tais propósitos estão relacionados com a capacidade dos dados de fomentar novas pesquisas e conhecimentos ao longo do tempo, garantindo a sustentabilidade e a interoperabilidade dos recursos informacionais. Para atingir tal objetivo, a curadoria digital busca agregar valor informacional aos dados por meio da representação da informação proveniente dos metadados. Sendo assim, o estudo questiona como os metadados influenciam as ações da curadoria digital e tem por objetivo analisar a relação entre os metadados e a curadoria digital. Trata-se de uma pesquisa de natureza teórica com abordagem qualitativa, realizada por meio de levantamento bibliográfico e da sistematização do estudo exploratório, recorrendo também a um estudo documental da plataforma *Europeana*, em razão de exemplificar a curadoria digital e a representação dos dados, no caso, os dados de patrimônios culturais disponibilizados no ambiente *Web*. Como resultado, verificou-se que a curadoria digital busca, por meio do processo de agregar valor informacional, enriquecer os dados de tal modo a facilitar o compartilhamento e a reutilização. Portanto, constatou-se que os metadados influenciam as ações da curadoria digital na medida em que esses elementos permitem explicitar as definições semânticas que são necessárias à comunicação dentro do processo de gerenciamento, principalmente de modo automatizado. Ao final, o estudo acerca da plataforma *Europeana* forneceu o embasamento para os resultados, demonstrando como os metadados, por meio de instrumentos, princípios, modelos e tecnologias, possibilitam que o contexto semântico dos recursos informacionais de patrimônios culturais seja assegurado. Conclui-se, por fim, que a importância dos metadados para a curadoria digital pode ser percebida não só como um fator essencial para garantir o planejamento e a execução de suas ações, mas principalmente para permitir que os recursos informacionais tenham seu propósito informacional assegurado e mantenham-se com qualidade e confiabilidade para o acesso, o uso e o reuso a longo prazo.

Palavras-chave: Curadoria Digital. Metadados. Representação da informação. *Web* Semântica. Patrimônios culturais.

ABSTRACT

The amount and variety of data associated with cultural heritage have grown significantly in digital environments, generating a diversity of possibilities and challenges related to information representation, especially in studies focused on digital curation. In the Web environment, there is need for data informational treatment to take into account, in addition to all the stages through which they circulate, the ability to guarantee their informational purposes. In this scenario, digital curation is highlighted as an interdisciplinary activity that seeks to establish guidelines not only to maintain and preserve informational resources but also to ensure that data are continuously consistent with their informational purposes. To achieve this goal, digital curation adds value to data through the representation of information from metadata. Thus, this study questions how metadata influence digital curation actions and aims analyze the relationship between metadata and digital curation. This theoretical qualitative study was carried out through bibliographic survey and exploratory systematization, also using a documentary study of the Europeana platform to exemplifying digital curation and data representation, in this case, cultural heritage data made available on the Web environment. Results showed that digital curation seeks, through the process of adding informational value, to enrich data in such a way as to facilitate sharing and reuse. Therefore, it was found that metadata influence digital curation actions to the extent that these elements enable explanation of the semantic definitions necessary for communication within the management process, mainly in an automated way. In the end, the study about the Europeana platform provided the basis for the results, demonstrating how metadata, through instruments, principles, models and technologies, allows the semantic context of informational resources of cultural heritage to be ensured. In conclusion, the importance of metadata for digital curation is perceived not only as an essential factor to guarantee planning and execution of its actions, but mainly to allow digital resources to have their informational purpose assured and remain with quality and reliability for long-term access, use and reuse.

Keywords: Digital Curation. Metadata. Information representation. Semantic Web. Cultural heritage.

LISTA DE ILUSTRAÇÕES

Figura 1 - “Bolo de noiva” da <i>Web Semântica</i>	44
Figura 2 - Tecnologias da <i>Web Semântica</i>	45
Figura 3 - Modelo funcional OAIS	60
Figura 4 - Modelo simplificado para o ambiente informacional JISC	62
Figura 5 - DCC Digital Curation Lifecycle Model	64
Figura 6 - Processos da curadoria digital do DCU	71
Figura 7 - <i>Extended Digital Curation Lifecycle Model (DCC&U)</i>	75
Figura 8 - Ciclo de vida dos recursos informacionais no ambiente digital	81
Figura 9 - Hierarquia de classes do EDM	102
Figura 10 - Representação em EDM da pintura Mona Lisa, de Leonardo da Vinci	103

LISTA DE QUADROS

Quadro 1 - Plano de Trabalho	19
Quadro 2 - Tipologia dos Padrões de Metadados	29
Quadro 3 - Tipos e funções dos Metadados	31
Quadro 4 - Características dos metadados	34
Quadro 5 - Princípios dos metadados.....	36
Quadro 6 - Definições de curadoria digital	54
Quadro 7 - Comparação entre os modelos de ciclo de vida de Curadoria	77
Quadro 8 - Metadados e suas funções para a curadoria digital	82
Quadro 9 - Princípios de Dados FAIR	84

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCSDS	<i>Consultative Comittee for Space Data Systems</i>
DCC	<i>Digital Curation Centre</i>
DCC&U	<i>Extended Digital Curation Lifecycle Model</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DCU	<i>Digital Curation Unit</i>
DigCCurr	<i>Digital Curation Curriculum</i>
EDM	<i>Europeana Data Model</i>
FAIR	<i>Findability, Accessibility, Interoperability e Reusability</i>
IDCC	<i>International Digital Curation Conference</i>
ISO	<i>International Organization for Standardization</i>
JISC	<i>Joint Information Systems Committee</i>
JSON	<i>JavaScript Object Notation</i>
LAM	<i>Libraries, Archives and Museums</i>
LISA	<i>Library and Information Science Abstracts</i>
LOD	<i>Linked Open Data</i>
MARC	<i>Machine Readable Cataloging</i>
NISO	<i>National Information Standards Organization</i>
OAIS	<i>Open Archival Information Systems</i>
OpenGLAM	<i>Open Galleries, Libraries, Archives and Museums</i>
OWL	<i>Ontology Web Language</i>
RDF	<i>Resource Description Framework</i>
RDFa	<i>Resource Description Framework in Attributes</i>
RIF	<i>Rule Interchange Format</i>
SciELO	<i>Scientific Electronic Library Online</i>
SKOS	<i>Simple Knowledge Organization System</i>
SPARQL	<i>Protocol and RDF Query Language</i>
TIC	Tecnologias da Informação e Comunicação
Turtle	<i>Terse RDF Triple Language</i>
URI	<i>Uniform Resource Identifier</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Problema de Pesquisa	15
1.2 Objetivos.....	15
1.3 Justificativa.....	16
1.4 Procedimentos metodológicos	19
1.5 Estrutura do trabalho	22
2 METADADOS E A REPRESENTAÇÃO DA INFORMAÇÃO EM AMBIENTES DIGITAIS	24
2.1 Tipos, funções e características dos metadados.....	27
2.2 Os metadados e as tecnologias semânticas.....	39
3 CURADORIA DIGITAL	48
3.1 Os modelos de curadoria digital	57
3.1.1 Modelo de referência OAIS.....	58
3.1.2 Modelo do JISC	61
3.1.3 Modelo de ciclo de vida de curadoria do DCC	63
3.1.4 Modelo de ciclo de vida de curadoria estendido - DCC&U.....	69
3.2 Análise e comparação dos modelos de curadoria digital.....	76
4 RELAÇÕES ENTRE METADADOS E CURADORIA DIGITAL	79
4.1 A contextualização semântica na curadoria digital	86
4.2 A curadoria digital na plataforma <i>Europeana</i>	94
5 CONSIDERAÇÕES FINAIS	110
REFERÊNCIAS	118

1 INTRODUÇÃO

Desde seu surgimento, a rede mundial de computadores, denominada *World Wide Web*, ou apenas *Web*, tem modificado as múltiplas dimensões da atuação humana, principalmente, nos setores relacionados com a produção e a disseminação de informação e de conhecimento, apoiados nas Tecnologias de Informação e Comunicação (TIC). Nesse cenário, a *Web* coloca-se como o principal meio pelo qual a informação é disponibilizada, tornando-se também a principal fonte de informação para diversos setores.

Nos espaços de circulação da informação, principalmente nas instituições que trabalham com patrimônios culturais, como bibliotecas, arquivos e museus, também denominadas como instituições LAM (acrônimo da expressão inglesa “*Libraries, Archives and Museums*”), as possibilidades trazidas pelos avanços tecnológicos resultaram em uma forte iniciativa de publicação de dados, na busca para ampliar as formas de compartilhamento e de uso para seus acervos e coleções. Marcondes (2016) destaca que os esforços em ampliar o acesso à informação e atender as demandas das sociedades modernas têm impulsionado a publicação e o compartilhamento de dados relativos aos patrimônios culturais.

Esses dados têm sido cada vez mais produzidos, disponibilizados e compartilhados de forma aberta e distribuída por instituições LAM, bem como por pessoas individuais (HYVÖNEN, 2012). Como define Hyvönen (2012, p. 1, tradução nossa) “Patrimônio cultural refere-se ao legado de objetos físicos, meio ambiente, tradições e conhecimento de uma sociedade, herdado do passado, mantido e desenvolvido no presente e preservado (conservado) para o benefício das gerações futuras”. Grande parte da memória social e também da história da humanidade podem ser apreendidas por meio desses legados, que constituem o domínio dos patrimônios culturais¹. Eles constituem importantes fontes de informação e de conhecimento e são cada vez mais encontrados em formatos digitais, a partir da publicação de seus dados.

O volume de dados vem se multiplicando à medida que o cenário tecnológico vigente permite maiores possibilidades de compartilhamento e de utilização de diversos tipos de conteúdo nos ambientes digitais. Tal cenário traz a possibilidade do desenvolvimento de serviços e espaços de interação digitais (SAYÃO, 2016), caracterizados por apoiar processos distribuídos e interoperáveis de criação e de compartilhamento de conteúdo, tais como os

¹ *Cultural heritage domain* é o termo comumente utilizado no âmbito internacional para designar o campo específico de conhecimentos e recursos informacionais que compõe a identidade sociocultural de uma comunidade e que devem ser preservados para assegurar sua perpetuidade.

grandes portais nacionais e internacionais de patrimônios culturais, como o *Europeana* e ainda grandes repositórios de dados abertos, como o *Linked Open Data Cloud* (HYVÖNEN, 2012). Como resultado, observa-se significativas mudanças na forma de trabalho das instituições que lidam com esses conteúdos, gerando desafios nos quais as abordagens tradicionais de tratamento informacional vêm tendo que ser repensadas.

Tais desafios encontram-se principalmente na constante tentativa de promover uma representação adequada dos recursos informacionais, que garanta sua recuperação, seu acesso, sua preservação, seu uso e reuso, ao mesmo tempo em que proporcione a interoperabilidade dos dados entre os diferentes recursos e os ambientes digitais (ALVES, 2016).

O tratamento da informação em meio digital abrange desde as formas de criação até sua divulgação, o que Glushko (2013) entende como uma manutenção de acervos ou curadoria. O termo ‘curadoria’ tem diversos ramos de desenvolvimento e pode ser encontrado em distintas áreas, se referindo a diferentes atividades. Em geral, seu conceito designa a ação de fazer e pensar em torno de algo que precisa de cuidados, caracterizando-se pelos procedimentos de gestão e administração².

A maior expressividade do uso da curadoria é no universo das instituições LAM e dos patrimônios culturais, estando ligada aos processos pelos quais os itens de uma coleção ou acervo são submetidos ao longo do tempo. Tais processos incluem desde ações para melhorar o acesso, incluindo possíveis intervenções, transformando a representação ou apresentação, até pesquisas para identificar, descrever e autenticar recursos informacionais (GLUSHKO, 2013).

Devido as suas características processuais, a curadoria pode ser associada à área de Ciência da Informação por designar os processos de representação, manutenção e conservação de recursos informacionais, visando à comunicação dos mesmos para fins de disseminação da informação. Portanto, no contexto dessa pesquisa, a palavra curadoria designa um conjunto de ações inter-relacionadas que visam à gestão e ao tratamento de qualquer conteúdo com valor informacional.

Em meio digital, a curadoria ganha novas contornos, não só dentro das instituições LAM, mas principalmente relacionando-se com o âmbito das TIC nos ambientes digitais, envolvendo práticas voltadas para a gestão de diversos tipos de recursos digitais por meio do

² Conceito extraído do relatório de iniciação científica intitulado “Análise de domínio e delineamento das atividades curatoriais para a Ciência da Informação”, de Autoria de Triques (2017), inserido no Programa Institucional de bolsas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na Universidade Federal de São Carlos (UFSCar) no processo: 119257/2016-5, com vigência de 01/08/2016 a 31/07/2017.

tratamento dos dados. Estes dados, por sua vez, podem ser provenientes de diversos contextos, desde gerados por pesquisas acadêmicas, até pelo registro de diversificados tipos de acervos, causando uma ampla variedade de dados com diferentes necessidades de tratamento informacional.

A denominada curadoria digital está ligada à contínua gestão dos dados, tendo como objetivo reduzir as ameaças ao valor informacional e também ao risco de obsolescência digital. Como esclarece Santos (2014 p. 130),

[...] a curadoria digital pode ser entendida como um termo guarda-chuva que está em franco desenvolvimento, que abarca definições correlatas voltadas à seleção, enriquecimento, tratamento e preservação da informação para o acesso e uso futuro, seja ela de natureza científica, administrativa ou pessoal.

Desse modo, ela visa garantir a manutenção de características como autenticidade, confiabilidade, integridade e usabilidade dos dados (HIGGINS, 2008). Entende-se que a manutenção de tais características está apoiada no tratamento proveniente da atribuição de metadados. Higgins (2007a) conceitua que os metadados constituem a ‘espinha dorsal’ da curadoria digital, ou seja, é o que permite que as ações realizadas sobre os recursos digitais se articulem e se efetivem. Os metadados, por sua vez, se configuram como informações descritivas ou contextuais que se referem ou estão associadas aos dados (HIGGINS, 2007a).

Os metadados são entendidos como dados que especificam as características de outros dados ou conjunto de dados, permitindo a representação dos recursos informacionais. Para os dados relativos aos patrimônios culturais, a preocupação em relação à representação proveniente da atribuição de metadados no contexto digital é de significativa importância, uma vez que tais dados são por natureza “[...] sintática e semanticamente heterogêneos, multilíngues, semanticamente ricos³ e altamente interligados” (HYVÖNEN, 2012, p. 8, tradução nossa). Sendo assim, os metadados garantem que um recurso informacional mantenha-se persistente em relação ao seu valor informacional.

Apesar da longa tradição em utilizar elementos descritivos para a organização e a representação das informações, o domínio dos patrimônios culturais tem se deparado nas últimas décadas com uma ampla variedade de possibilidades e de desafios relacionados à representação da informação, envolvendo não só a descrição com vistas à recuperação, mas também aos aspectos voltados à sustentabilidade e à interoperabilidade nos ambientes digitais.

³ *Semantically Rich* é o termo usado por Hyvönen (2012) para indicar um contexto com complexas conexões conceituais.

1.1 Problema de Pesquisa

Com o volume e a variedade de dados cada vez maior nos ambientes digitais, amplia-se a preocupação não só com o acesso e recuperação da informação, mas também com suas implicações a longo prazo. Diante disso, são exigidas novas formas de tratamento informacional que levem em consideração todos os estágios pelos quais os dados circulam e ainda, a capacidade de garantir o propósito informacional a que um recurso se destina.

Na medida em que os sistemas informacionais se tornam cada vez mais orientados ao processamento e a troca de informações semanticamente, cresce a importância em garantir esse propósito informacional. Assim, recorre-se à curadoria digital como uma possível solução para os desafios de gerenciamento de dados, tendo em vista que seu objetivo envolve não só a manutenção e a preservação dos recursos informacionais, mas principalmente agregar valor informacional aos dados para que esses recursos possam atender aos diferentes propósitos a que são destinados nos ambientes em que circulam.

A curadoria digital leva em consideração a representação não só com vistas à recuperação, mas também considera a sustentabilidade e a interoperabilidade dos dados. Para tanto, os metadados são elementos centrais, possibilitando assegurar a autenticidade, a confiabilidade, a integridade e a usabilidade dos dados para a adequação ao seu propósito informacional. No entanto, é importante considerar que a garantia de autenticidade, a confiabilidade, a integridade e a usabilidade dos dados não são uma tarefa simples, implicando em adicionar diversas camadas de metadados com diferentes tipologias que acabam por potencializar os desafios da curadoria digital.

Nessa perspectiva, busca-se o enfoque nas ações da curadoria digital dirigidas pela utilização dos metadados, questionando **como os metadados influenciam as ações da curadoria digital?** A resposta para esta questão reside na dimensão relacional entre a curadoria digital e a representação da informação proveniente dos metadados.

1.2 Objetivos

Como objetivo geral propõe-se analisar a relação entre os metadados e a curadoria digital, tendo em vista a representação dos dados no ambiente *Web*.

Dessa forma, os objetivos específicos são:

- Descrever os fundamentos da curadoria digital e dos metadados no âmbito da representação da informação no ambiente *Web*;
- Analisar os modelos de curadoria digital, tendo em vista suas ações e objetivos;
- Verificar como os metadados possibilitam assegurar o contexto semântico dos recursos informacionais;
- Apresentar a *Europeana* como uma plataforma digital de representação de dados no ambiente *Web*.

1.3 Justificativa

Diferente do paradigma que deu início a consolidação da Ciência da Informação na década de 1960, com a introdução da tecnologia da computação na automação de tarefas como descrição e recuperação da informação, a Ciência da Informação enfrenta outro paradigma, no qual não apenas a descrição das informações se manifesta em formatos digitais, mas também as próprias informações (HJØRLAND, 2014). Esse cenário é marcado pelas múltiplas formas de divulgação e de disponibilização das informações, principalmente por meio do ambiente *Web*.

O ambiente *Web* ganhou importância significativa na produção e na disseminação de informação e de conhecimento, implicando diretamente no modo de atuação de diversos setores, em especial no domínio dos patrimônios culturais, representados por instituições como bibliotecas, arquivos e museus, também denominadas como instituições LAM.

Diante das possibilidades de organizar, de compartilhar e de processar as informações sobre recursos de patrimônios culturais em meio digital, destaca-se a necessidade de refletir no âmbito da Ciência da Informação, os procedimentos e os métodos envolvidos no tratamento informacional desses dados, em especial, os dados que estão alocados nos ambientes informacionais direcionados pela construção semântica e pela ligação entre os dados, como na *Web*.

Nesse contexto, a pesquisa se enquadra na linha 2 do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de São Carlos, denominada “*Tecnologia, Informação e Representação*”, a qual contextualiza-se diante da crescente influência dos aparatos tecnológicos no modo de atuação da sociedade, afetando, consequentemente, os

processos de geração, de armazenamento, de representação, de organização, de compartilhamento, de disseminação, de uso, de reuso e de preservação de informações em ambientes informacionais. Sendo assim, a pesquisa se insere na dimensão teórica que é fundamentada no uso estratégico das Tecnologias da Informação e Comunicação (TIC) na área de Ciência da Informação, buscando com os resultados otimizar a inovação na área de Organização e Representação da Informação e do Conhecimento.

Para tanto, destaca-se a curadoria digital, como um campo de estudo interdisciplinar que se dedica a criar estratégias e procedimentos visando além da preservação e do acesso contínuo dos dados em ambientes digitais, também a garantia de sua autenticidade, confiabilidade, integridade e usabilidade para a adequação ao seu propósito informacional e contexto semântico.

A perspectiva acerca da curadoria digital considerada nesse estudo parte da pesquisa de Constantopoulos e Dallas (2008), que tem por base o extenso trabalho sobre dados de patrimônios culturais em meio digital do *Digital Curation Unit* (DCU), no qual defendem que a curadoria digital precisa de uma abordagem que assegure uma representação adequada e acesso de longo prazo aos dados, à medida que seu contexto de uso muda.

Isso se torna fundamental diante do volume de dados que são continuamente disponibilizados no ambiente *Web*, e principalmente, devido ao fato de que em meio aberto e distribuído, as demandas dos usuários não são previsíveis ou pré-estabelecidas e os recursos podem ser utilizados para vários propósitos e de diferentes formas.

Esses usos e propósitos estão relacionados com a capacidade desses recursos informacionais servirem como base para novas pesquisas e gerarem novos conhecimentos. No caso dos recursos informacionais de patrimônios culturais, devido à sua natureza específica, seus dados são caracteristicamente heterogêneos, semanticamente ricos, multilíngues e altamente interligados. Tal riqueza e variedade de conteúdo traz naturalmente desafios para o tratamento informacional.

Nessa perspectiva, discute-se a relação da curadoria digital com a representação proveniente dos metadados, tendo em vista que “[...] uma parte fundamental do processo de curadoria é garantir que os metadados estejam disponíveis para descrever os conjuntos de dados para uso futuro” (CHAO, 2014, não paginado, tradução nossa). Ao mesmo tempo, compreende-se a importância dos metadados para os objetivos da curadoria digital, considerando que apesar desses elementos serem essenciais, são ao mesmo tempo o fator que mais restringe a sua eficácia, pois

[...] há limites para os recursos humanos que podem ser aplicados à curadoria, dadas as enormes quantidades de dados envolvidos, de modo que a automação do maior número possível de processos de curadoria é necessária. Metadados são críticos para suportar esse alto nível de automação (HARVEY, 2010, p.79, tradução nossa).

Considerando esse ponto de vista, verifica-se uma deficiência na literatura, principalmente no âmbito nacional, a respeito de estudos que relacionem as implicações desse novo cenário de automação dos processos de tratamento informacional para a curadoria digital. Principalmente no que diz respeito às possibilidades de organização e de representação da informação e do conhecimento apresentadas pelas tecnologias da *Web Semânticas* e sua influência na gestão de dados de patrimônios culturais. Ademais, foi possível constatar que os estudos a respeito do modelo *Extended Digital Curation Lifecycle Model* (DCC&U) (CONSTANTOPOULOS; DALLAS, 2008, CONSTANTOPOULOS; DALLAS *et al.*, 2009), que introduziu a ideia de uma camada adicional na curadoria digital para a representação dos aspectos do contexto semântico dos recursos informacionais, não tiveram mais progressos significativos.

Sendo assim, a pesquisa se justifica pela crescente popularização da temática que envolve a curadoria digital, ao passo que ainda é um campo de estudo em desenvolvimento que carece de fundamentação teórica, principalmente em relação a sua aplicação prática em ambientes informacionais na *Web*, orientados pelas tecnologias da *Web Semântica*.

Para a área da Ciência da Informação, a pesquisa busca trazer contribuições tanto por meio de reflexões teóricas sobre a curadoria digital, quanto por meio do esclarecimento do papel dos metadados no gerenciamento de dados em ambientes informacionais digitais. Além disso, pauta-se na preocupação relacionada ao tratamento informacional do crescente volume de dados que são provenientes tanto da publicação, quanto da criação em larga escala.

Como relevância social, essa pesquisa se justifica por buscar demonstrar à comunidade científica e aos usuários de patrimônios culturais, como o emprego de metadados de qualidade no processo de curadoria digital pode proporcionar maiores e melhores possibilidades de acesso ao conteúdo disponibilizado em ambientes digitais, garantindo uma melhor experiência aos usuários e fomentando novos conhecimentos.

Ademais, a pesquisa tem motivação pessoal, sendo justificada pelo interesse no aprofundamento dos estudos da Iniciação Científica, fomentada pelo Conselho Nacional de

Desenvolvimento Científico e Tecnológico (CNPq)⁴ e do Trabalho de Conclusão de Curso (TCC) de Triques (2017).

1.4 Procedimentos metodológicos

A pesquisa configura-se como bibliográfica, de abordagem qualitativa, de caráter exploratório e de natureza teórica, recorrendo a um estudo documental da plataforma *Europeana* para embasamento dos resultados. Possui a análise exploratória pautada em um levantamento sistemático e qualitativo dos dados, sendo estes analisados de modo indutivo.

Por meio da abordagem indutiva, cuja característica é uma “[...] análise que parte do particular e coloca a generalização como um produto posterior do trabalho de coleta de dados particulares” (GIL, 2008, p.10), buscou-se observar, a partir da análise do conteúdo das publicações, a relação entre as ações que caracterizam a curadoria digital e a influência dos metadados nas mesmas. Posteriormente, comparou-se essas publicações com a finalidade de descobrir as relações existentes entre elas. Por fim, com base na relação verificada, identificou-se uma situação concreta na qual os resultados observados pudessem ser analisados na realidade.

Em relação ao levantamento sistemático, este caracteriza-se pela observação sistemática da recuperação e análise crítica da literatura, possibilitando a delimitação dos aspectos a serem considerados para alcançar os objetivos pretendidos (GIL, 2008). Sendo assim, para a execução da pesquisa, foi desenvolvido um plano de trabalho, apresentado no quadro 1, no qual buscou-se contemplar as principais questões estabelecidas nos objetivos deste trabalho, para que pudessem ser localizadas as contribuições científicas pertinentes ao tema.

Quadro 1 - Plano de Trabalho

ETAPAS	ATIVIDADES
<p>1ª etapa: Levantamento bibliográfico e seleção do material</p>	<p>Identificação das publicações para criar embasamento teórico sobre a temática, por meio do levantamento bibliográfico realizado em nível nacional e internacional.</p>

⁴ Projeto 800675/2016-9, intitulado como “Análise de domínio e delineamento das atividades curatoriais para a Ciência da Informação”, inserido no Programa Institucional de bolsas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) na Universidade Federal de São Carlos (UFSCar) no processo: 119257/2016-5, com vigência de 01/08/2016 a 31/07/2017.

<p>2ª etapa: Leitura e interpretação</p>	<p>Leitura do material selecionado para o desenvolvimento da base teórica para a comparação dos diferentes pontos de vista identificados na literatura sobre o tema.</p>
<p>3ª etapa: Análise e estabelecimento das características fundamentais extraídas da literatura</p>	<p>Análise das principais características encontradas na literatura para elucidação do problema de pesquisa.</p>
<p>4ª etapa: Sistematização do estudo exploratório</p>	<p>Identificação das relações existentes entre os fundamentos selecionados, de forma a elucidar tal relação à luz da fundamentação teórica.</p>
<p>5ª etapa: Elaboração da redação final e Divulgação da pesquisa</p>	<p>Desenvolvimento do relatório textual para divulgação à comunidade científica dos resultados obtidos.</p>

Fonte: Elaborado pela autora.

Na primeira etapa, foi realizado o levantamento bibliográfico em nível nacional e internacional para identificação e seleção do material que desse embasamento teórico e direcionamentos sobre a temática. Para isso, foram utilizadas fontes bibliográficas como fundamentação. Gil (2008, p. 44) conceitua que “[...] a pesquisa bibliográfica é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos”. Além disso, a pesquisa tem caráter exploratório, com a finalidade de proporcionar a familiaridade com a área de estudo e a sua delimitação (GIL, 2008), fornecendo base teórica para a construção de um conhecimento científico.

A seleção das publicações obedeceu aos critérios de pertinência com relação aos objetivos específicos. Esses critérios referem-se a delimitação das publicações pertinentes ao tema no âmbito da Ciência da Informação nos idiomas português, inglês e espanhol, com período de publicação a partir do ano 2000, ocasião em que se observa o início do crescimento das discussões acerca da temática.

Dessa forma, foram consultadas fontes primárias, secundárias e terciárias nas bases de dados especializadas da área de Ciência da Informação, tais como Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI) e *Library and Information Science Abstracts* (LISA), bem como periódicos da área de Ciência da Informação no Portal de Periódicos da CAPES. Além disso, foram consultadas as bases internacionais *Web of Science* e *Scientific Electronic Library Online* (SciELO), bem como *Google Scholar*.

Também foram incluídas na bibliografia pesquisada publicações do *The International Journal of Digital Curation*, periódico de referência na temática da curadoria digital.

Além das fontes citadas, foram utilizados autores para a fundamentação acerca da Organização e da Representação da Informação e do Conhecimento e *Web Semântica*, como Berners-Lee, Hendler e Lassila (2001), Méndez Rodríguez (2002), Senso e Rosa Piñero (2003), Souza e Alvarenga (2004), Alves (2010), Santos e Sant’Ana (2013), Araújo (2014), Pomerantz (2015), Zeng e Qin (2008; 2016), Gilliland (2008; 2016), Gartner (2016) e Riley (2004; 2017).

Para contemplar as publicações pertinentes ao tema, as estratégias de busca que foram utilizadas envolvem: ‘curadoria digital’, ‘modelo de ciclo de vida de curadoria digital’, ‘metadados’, em português, ‘*digital curation*’, ‘*curation lifecycle model*’, ‘*metadata*’, em inglês, e ‘*curaduría digital*’, ‘*modelo de ciclo de vida de curaduría digital*’, ‘*metadatos*’, em espanhol. Também foram usados os termos ‘*web semântica*’, ‘*curadoria de dados*’ e ‘*patrimônios, culturais*’, em português, bem como os termos ‘*semantic web*’, ‘*data curation*’ e ‘*cultural heritage*’ em inglês e ‘*web semántica*’, ‘*curaduría de datos*’ e ‘*patrimonio cultural*’ em espanhol.

Na segunda etapa da pesquisa, realizou-se a leitura e a interpretação do material selecionado para o estabelecimento da base teórica da pesquisa, contemplando os dois primeiros objetivos específicos. Estabelecida a base teórica, a terceira etapa foi analisar as características fundamentais extraídas da literatura para elucidar o problema de pesquisa, o que oportunizou contemplar o terceiro e o quarto objetivos específicos. Na quarta etapa foram comparadas e discutidas as características fundamentais extraídas da literatura, abrangendo assim o objetivo geral dessa pesquisa. A quinta e última etapa, consistiu no desenvolvimento do relatório textual para divulgação dos resultados obtidos visando à comunicação científica.

Sendo assim, o desenvolvimento da pesquisa foi pautado em uma análise qualitativa do conteúdo dos dados obtidos, identificando e comparando as abordagens das publicações e a comunidade de autores. Desse forma, atendendo ao plano de trabalho e aos critérios propostos, os resultados foram analisados da seguinte forma:

A partir das etapas de sistematização e análise do conteúdo das publicações científicas selecionadas, buscou-se verificar como os metadados influenciam as ações da curadoria digital, principalmente no que diz respeito ao objetivo de manter o propósito informacional dos recursos digitais em diversos contextos de uso, tendo em vista o cenário tecnológico vigente, caracterizado pelo alto nível de automação e pela troca e processamento de informações semanticamente.

Para isso, realizou-se em um primeiro momento uma investigação teórica acerca dos fundamentos que regem os dois objetos de pesquisa definidos nesse trabalho: os metadados e a curadoria digital. Para tanto, foi realizada uma revisão de literatura que contempla os metadados em ambientes digitais, buscando conceituá-los diante das dimensões do tratamento informacional trazidas pelas tecnologias desenvolvidas para o ambiente *Web*. Em seguida, também foram investigados os fundamentos da curadoria digital, tendo como base o modelo elaborado pelo *Digital Curation Centre* (DCC) e buscando identificar demais modelos, de modo a elucidar suas práticas e objetivos.

Os modelos apresentados foram selecionados por meio da revisão da literatura científica, método adotado na pesquisa, tendo como critério de seleção o fato destes serem aplicáveis à diversos contextos e de não serem restritivos a coleções e sistemas específicos. Tal escolha foi adotada para que esses modelos apresentassem uma panorama da curadoria digital de forma abrangente de modo a identificar suas principais características, contribuindo assim para a elucidação da discussão proposta. A análise dos modelos partiu de caracterização de cada um, buscando identificar as similaridades e as diferenças entre eles, com base na literatura científica.

Tendo essas características e fundamentos acerca da curadoria digital e dos metadados identificados, a investigação foi direcionada em um segundo momento para verificar quais seriam as implicações dos metadados junto às ações da curadoria digital. Com base na revisão de literatura e nos resultados obtidos com a comparação dos modelos de curadoria digital, foi possível identificar os pontos de intersecção que associam os metadados à prática da curadoria, oportunizando a discussão acerca da sinergia entre a curadoria digital e a Ciência da Informação e, permitindo ainda, um aprofundamento teórico sobre a representação da informação no ambiente digital e sua relação com os aspectos semânticos e a ligação entre dados.

1.5 Estrutura do trabalho

Os temas apresentados e discutidos ao longo do trabalho estão distribuídos em seções e subseções. A seção inicial delimita o tema e o problema de pesquisa, apresentando objetivos, justificativa, procedimentos metodológicos e cronograma da pesquisa. As demais seções são apresentadas a seguir:

SEÇÃO 2 - METADADOS E A REPRESENTAÇÃO DA INFORMAÇÃO EM AMBIENTES DIGITAIS: Apresenta uma revisão de literatura a respeito da representação da

informação em ambientes digitais no âmbito da Ciência da Informação, apresentando os tipos, as funções e as características dos metadados, bem como um breve panorama de seu uso no ambiente *Web*.

SEÇÃO 3 - CURADORIA DIGITAL: Apresenta uma revisão de literatura a respeito da curadoria digital, abordando os principais aspectos relacionados ao seu surgimento, desenvolvimento e características, bem como suas ações e objetivos que a tornam uma atividade de grande potencial para o gerenciamento de dados. Nessa seção são apresentados, analisados e comparados os modelos de curadoria digital selecionados para o estudo.

SEÇÃO 4 - RELAÇÕES ENTRE METADADOS E CURADORIA DIGITAL: Discute a importância da representação da informação para a curadoria digital, buscando contemplar o objetivo geral da pesquisa ao apresentar as implicações dos metadados para o gerenciamento dos dados, em especial no ambiente *Web*, no qual a atribuição de metadados é essencial para o efetivo funcionamento dos sistemas.

SEÇÃO 5 - CONSIDERAÇÕES FINAIS: Aborda as considerações finais do estudo, delineando os resultados obtidos a partir da revisão bibliográfica e das discussões desenvolvidas, fornecendo ainda direcionamentos para estudo futuros.

2 METADADOS E A REPRESENTAÇÃO DA INFORMAÇÃO EM AMBIENTES DIGITAIS

Esta seção dedica-se a analisar a literatura a respeito da representação da informação no âmbito da Ciência da Informação, apresentando os tipos, as funções e as características dos metadados, bem como um breve panorama de seu uso no ambiente *Web*, buscando assim discutir temas pertinentes à problemática proposta. Desse modo, inicia-se esta seção a partir dos conceitos de dados e de metadados.

Os dados podem ser compreendidos como uma expressão mínima de conteúdo e podem ser produzidos de inúmeras formas, seja por humanos ou por máquinas, eles estão presentes em todas as possibilidades de produção de recursos informacionais, bem como nas características dos diversos tipos de recursos existentes.

Santos e Sant'Ana (2013, p. 205) definem que dado é

[...] uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor, de tal forma que, mesmo que não esteja explícito o detalhamento sobre contexto do conteúdo, ele deverá estar disponível de modo implícito no utilizador, permitindo, portanto, sua plena interpretação.

Os dados também assumem a função de especificar características de outros dados, descrevendo a forma como eles serão utilizados, exibidos e contextualizados (SOUZA; ALVARENGA, 2004). Nesse caso, esses tipos de dados são denominados como metadados.

Em sistemas automatizados, a interpretação de dados requer a utilização de dados que descrevam outros dados. O termo 'metadados', que significa literalmente 'dados sobre dados', foi cunhado no final da década de 1960 para se referir a um conjunto de declarações sobre os dados (POMERANTZ, 2015). Sendo um termo amplamente utilizado, ainda que com diferenças de terminologia, os metadados fazem parte do cotidiano de diversas comunidades profissionais que projetam, criam, descrevem, preservam e usam sistemas e recursos informacionais (GILLILAND, 2016).

Como destacam Zeng e Qin (2016, p.12-13, tradução nossa), diferentes comunidades podem se referir aos mesmos componentes de metadados com terminologia diferente:

Por exemplo, nos metadados da ciência da computação existem informações sobre objetos de banco de dados e/ou objetos de programa, como tabelas e procedimentos armazenados. O designer de banco de dados pode chamar elementos de metadados "campos de dados" em vez de "elementos". Outros termos como "propriedades" e "atributos" também são frequentemente usados na literatura de metadados, mas seus significados mudam o contexto no qual esses termos são usados.

Embora o termo ‘metadados’ seja usado há apenas algumas décadas, só popularizando-se entre as décadas de 80 e 90, com o advento dos sistemas de gerenciamento de bancos de dados e com a criação do padrão *Dublin Core*, o conceito por trás do nome já era há muito tempo utilizado (ALVES, 2010; POMERANTZ, 2015).

O contexto dos patrimônios culturais, tem uma longa história de criação e compartilhamento de metadados (RILEY, 2017). Ainda que tenha sido amplamente difundido para designar representação informacional em meio digital (ALVES, 2010), o termo metadados pode ser vinculado a uma conceituação que independe do ambiente digital, se referindo ao tratamento informacional proveniente da representação da informação. Como explica Alves (2010), os metadados remontam a representação descritiva já realizada no passado, configurando-se como uma denominação atual no processo de tratamento da informação, se referindo a elementos que permitem a construção de representações para um recurso informacional.

Para a comunidade de trabalho com os patrimônios culturais, o termo metadados é comumente aplicado “[...] às informações de valor agregado que criam para organizar, descrever, rastrear e melhorar o acesso a objetos de informação e itens físicos e coleções, relacionados a esses objetos”. (GILLILAND, 2016, p. 2, tradução nossa). Por isso, os metadados são considerados essenciais nos processos de representação, seja ele em quaisquer ambientes onde a informação circula. Eles podem ser definidos como:

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação. (ALVES, 2010, p. 47).

Todos os recursos informacionais possuem declarações que podem ser feitas sobre eles. Independentemente do tipo ou da forma do recurso, tais declarações referem-se ao conteúdo, contexto ou estrutura. O conteúdo está relacionado ao que o recurso apresenta, de forma intrínseca; o contexto indica como o recurso se apresenta, da criação à vida útil, isto é, a sua forma extrínseca; e por fim, a estrutura refere-se às associações do recurso ou entre os recursos informacionais, podendo ser intrínseca ou extrínseca, ou ainda, ambas (GILLILAND, 2016). Desse modo, tais declarações viabilizam a identificação e o entendimento acerca do recurso informacional, configurando-se como elementos de representação ou metadados.

Como salienta Marcondes (2001, p. 61) “De nada adianta a informação existir, se quem dela necessita não saber da sua existência, ou se ela não puder ser encontrada”. Há tempos essa questão já é levada em consideração por setores tradicionais como as bibliotecas e demais instituições, que sempre fizeram uso das tecnologias vigentes para o desenvolvimento e o aprimoramento de métodos e de instrumentos para a representação visando à recuperação da informação (ALVES, 2010).

Para que haja recuperação da informação é necessário que os aspectos elementares dessa informação e de sua estrutura de armazenamento sejam explicitados de alguma forma. Por isso a representação da informação é uma disciplina fundamental para o tratamento informacional, pois é a partir dela que se oferece as mínimas condições para identificar e recuperar um recurso informacional.

Durante o processo de representação, o recurso passa por um processo de condensação e de sintetização de sua apresentação. A apresentação de um recurso informacional, pode-se dizer que é a forma que o objeto se forma no mundo real, com os mesmos valores físicos e de seu conteúdo. Por meio de uma representação, esses valores transformam-se em sucintas estruturas que auxiliam no acesso e na localização (SIMIONATO, 2015).

Assim, a representação pode ser entendida como um processo cognitivo que é primeiramente inerente à natureza humana, e que em um segundo momento passa a ser social, implicando além da concretização e da fixação do conhecimento em determinada estrutura, também a comunicação e o compartilhamento de ideias (BAPTISTA, 2007). É o que Marcondes (2001) identifica como uma segunda natureza humana, que é chamada de cultura e se caracteriza como intersubjetiva e capaz de ultrapassar os limites de espaço e de tempo.

Baptista (2007) aponta que a representação pode ser categorizada em diferentes níveis, sendo a representação primária aquela que vai do pensamento para o objeto; a representação secundária, a que vai do objeto para o registro; e por fim, a representação terciária, que vai do primeiro registro para o segundo registro.

Essa categorização permite vislumbrar a estreita relação do processo de representação da informação com a Ciência da Informação, “[...] na medida em que o objeto precisa ser não somente representado como também transformado em recurso informacional, como elemento constitutivo básico do conhecimento” (BAPTISTA, 2007, p. 180). Isso significa que representar um recurso lhe dá condições de manter seu potencial informacional para transmitir e gerar conhecimento.

Assim, destaca-se que criar representações é algo fundamental e inerente à natureza humana e que, portanto, tem grande interesse para a Ciência da Informação, na medida em que possibilita comunicar, registrar e perpetuar o conhecimento humano.

2.1 Tipos, funções e característica dos metadados

Tendo em vista que os registros do conhecimento humano são cada vez mais encontrados e mantidos em formato digital, constituindo os recursos informacionais digitais, destaca-se a importante função dos metadados como elementos constitutivos básicos para a representação da informação e do conhecimento também em ambientes digitais. Para os ambientes digitais, os metadados proporcionam a base da estrutura de descrição dos recursos informacionais (SANTOS; ALVES, 2009) e vão além, estando presentes em todas as etapas do ciclo de vida dos dados.

Nesses ambientes, caracterizados pela abundância de informação heterogênea e multilíngue, nos quais os problemas tradicionais de representação da informação e do conhecimento, bem como de recuperação, são ampliados, os metadados se tornam elementos centrais para gerenciar, preservar, estruturar, disponibilizar e conectar dados.

Sayão (2012) aponta que é impossível discutir serviços e sistemas de informação sem o envolvimento direto com questões relacionadas aos metadados.

Na medida em que a ideia de metadados se torna uma parte essencial do mundo digital eles se mostram conceitualmente mais complexos e mais abrangentes, apoiando um espectro extremamente amplo de atividades. Essa novas dimensões de metadados são vitais para o acesso e para a interpretação dos recursos informacionais digitais como são importantes também para a estruturação e para os processos de gestão associados a esses recursos, que podem incluir inúmeras funções, tais como: controle dos direitos, intercâmbio, comércio eletrônico, interoperabilidade técnica e semântica, reuso da informação e curadoria digital [...] (SAYÃO, 2012, p. 109-110).

Com isso, pode-se fazer um paralelo com os papéis ou as funções que os metadados vêm assumindo. Além dos papéis tradicionais já desempenhados por esses elementos, como: identificação e descrição da informação; busca e recuperação; e localização dos documentos, os metadados ainda passam a ter papéis emergentes, resultantes das características dos ambientes digitais, como: autoria e propriedade intelectual; formas de acesso; atualização da informação; preservação e conservação; restrição de uso; valoração do conteúdo; visibilidade da informação; acessibilidade dos conteúdos (MÉNDEZ RODRÍGUEZ, 2002).

Os metadados estão, portanto, vinculados a uma grande variedade de funções que incluem desde as informações descritivas, envolvendo a recuperação e a descoberta de recursos informacionais, até as informações relativas à gestão, ao uso e à preservação dos recursos informacionais, que referendam localização, formato, validação, processamento etc.

Para serem mais efetivos nessas funções, os metadados podem apresenta-se em uma estrutura padronizada, compondo um padrão ou esquema de metadados. Esses padrões ou esquemas suportam uma série de funções definidas, as quais permitem especificar elementos que tornam a representação possível, descrevendo o recurso informacional e viabilizando a identificação, a localização e a recuperação e, ao mesmo tempo, facilitando o gerenciamento e o acesso (HIGGINS, 2007a).

Os padrões de metadados podem ser conceituados como:

[...] padrões que estabelecem regras para a definição de atributos (metadados) de recursos de informacionais, para a) obter coerência interna entre os elementos por meio de semântica e sintaxe; b) promover necessária facilidade para esses recursos serem recuperados pelos usuários; c) permitir a interoperabilidade dos recursos de informação. (ALVES, 2010, p. 47).

Tal como acontece em relação à descrição dos recursos informacionais analógicos, o uso de esquemas de codificação apresenta-se como um elemento-chave para a construção padronizada de metadados em grande escala, como ocorre em ambientes digitais, viabilizando o compartilhamento e a recuperação da informação.

Por essa razão,

[...] os metadados e padrões de metadados continuam a desempenhar um papel fundamental para a representação adequada dos recursos informacionais, proporcionando não só a recuperação do patrimônio cultural em ambiente Web, mas também seu acesso e sua preservação em longo prazo, permitindo a interoperabilidade dos dados entre ambientes de diferentes áreas (ALVES, 2018, p.97).

Percebe-se que, assim, a busca por um tratamento informacional eficaz e por uma padronização consistente, ainda se perpetua, porém o que vêm mudando são as perspectivas desse tratamento, que sofrem influência dos avanços tecnológicos.

Higgins (2007a) destaca que usar padrões de metadados, desde o início de um projeto, garante metadados de qualidade e consistentes para suportar a descoberta, o uso e a integridade de recursos digitais a longo prazo. Além disso, é importante considerar que para uma efetiva implementação de padrões de metadados, os tipos de padrões existentes em relação à estrutura, ao conteúdo e à funcionalidade, devem ser analisados, pois diferentes

tipos de padrões de metadados podem ser usados de forma interdependente para atingir diversos objetivos (HIGGINS, 2007b).

Gilliland (2016) faz, nessa perspectiva, uma divisão segundo a tipologia de padrões (Quadro 2), organizando-os em categorias e fornecendo exemplos de uma variedade de esquemas e de formatos de metadados que permitem estruturar e expressar de maneiras padronizadas os dados para processamento, publicação e implementação em máquinas.

Quadro 2 - Tipologia dos Padrões de Metadados

TIPO	EXEMPLO
Padrões de estrutura de dados (conjuntos de elementos de metadados, esquemas).	<i>MARC (Machine-Readable Cataloging) Format, Encoded Archival Description (EAD), BIBFRAME (Bibliographic Framework), Dublin Core Metadata Element Set, Categories for the Description of Works of Art, VRA Core.</i>
Padrões de valor de dados (vocabulários controlados, tesouros, listas controladas).	<i>Library of Congress Subject Headings, Name Authority File, and Thesaurus for Graphic Materials; Getty Art & Architecture Thesaurus, Union List of Artist Names (ULAN), and Thesaurus of Geographic Names; ICONCLASS; Medical Subject Headings.</i>
Padrões de conteúdo de dados (regras de catalogação e códigos).	<i>Anglo-American Cataloguing Rules, Resource Description and Access, International Standard Bibliographic Description, Cataloging Cultural Objects, Describing Archives: A Content Standard.</i>
Padrões de formato de dados/intercâmbio técnico (padrões de metadados expressos em formato legível por máquina).	<i>Resource Description Framework, MARC21, MARCXML, EAD XML DTD, METS, BIBFRAME, LIDO XML, Simple Dublin Core XML, Qualified Dublin Core XML, VRA Core 4.0 XML.</i>

Fonte: Adaptado de Gilliland (2016).

A divisão é feita por padrões em estrutura de dados, que contemplam os conjuntos de elementos de metadados e esquemas que compõem um registro ou outras informações sobre o objeto. Os padrões de valor de dados são os vocabulários controlados, tesouros, listas de controle, ou seja, são os termos e outros valores utilizados para preencher os conjuntos de elementos de metadados. Já os padrões de conteúdo de dados são orientações para o formato e a sintaxe dos valores de dados que são utilizados para preencher elementos de metadados. São consideradas também as regras e os códigos para a descrição. Por fim, os padrões de

intercâmbio são uma forma de codificação e de estrutura que permite expressar os dados em uma linguagem de marcação processável por máquina (GILLILAND, 2016).

Os metadados padronizados permitem a recuperação automática da informação, o que promove a consistência dos bancos de dados e viabiliza o compartilhamento de informações entre eles, ou seja, a interoperabilidade e o intercâmbio de dados (CASTRO; SIMIONATO; ZAFALON, 2016; ZENG; QIN, 2016; RILEY, 2017). Por sua vez, Higgins (2007a) destaca que os padrões de metadados são compostos por vários elementos que podem ser categorizados nas diferentes funções que eles suportam, podendo ser divididos quanto à finalidade de seu uso em diferentes tipos.

Diversos autores buscam categorizar os metadados por seus tipos e funções, entre eles Méndez Rodríguez (2002), Senso e Rosa Piñero (2003), Alves (2010), Pomerantz (2015), Zeng e Qin (2008; 2016), Gilliland (2008; 2016), Gartner (2016) e Riley (2004; 2017). A análise comparativa da literatura permite constatar que há certa divergência em relação a essa categorização. No entanto, é possível verificar que, apesar das distinções, a divisão por categorias auxilia no entendimento acerca do papel que determinado tipo de metadados tem em um sistema de informação.

Para alguns autores, os metadados podem ser divididos em cinco tipos: **administrativos**, que são os metadados usados no gerenciamento e na administração dos recursos informacionais; **descritivos**, que se referem aos metadados usados para descrever e identificar informações sobre recursos; **preservação/conservação**, que são os metadados relacionados à conservação de recursos de informação; **técnicos**, que são os metadados relacionados ao funcionamento do sistema e comportamento dos próprios metadados; e **uso**, que se referem aos metadados relacionados com o nível e o tipo de uso dos recursos informacionais (GILLILAND, 2008; 2016; SENSO ; ROSA PIÑEROS, 2003; ALVES, 2010; ZENG; QIN, 2008; 2016).

Para Pomerantz (2015), três grandes categorias de metadados são consideradas, sendo elas: metadados **descritivos**, **administrativos** e de **uso**. Novamente os metadados descritivos se encarregam das declarações que podem ser feitas sobre os recursos para sua identificação e localização. Já os metadados administrativos, são também considerados uma categoria guarda-chuva, no qual encontra-se subtipos como os metadados técnicos, estruturais, de proveniência, de preservação, de direitos e *meta-metadata*. E por fim, os metadados de uso, que envolvem dois subtipos de metadados, que Pomerantz (2015) refere-se como “*data exhaust*” e “*paradata*”. O primeiro diz respeito às trilhas dos dados, ou seja, o histórico de navegação que os usuários deixam ao realizar suas atividades na *Web* e o segundo refere-se

aos dados que informam sobre como, por que e por quem os recursos estão sendo usados, isto é, o histórico de uso.

Já para os autores Riley (2004) e Gartner (2016), a categorização difere um pouco das outras, estabelecendo três categorias que são: **metadados descritivos**, que se referem aos metadados que auxiliam descobrir e localizar os dados a que são associados; **metadados estruturais**, que se destinam a definir as estruturas que reúnem as parte de um recurso em algo mais complexo; e **metadados administrativos**, que se referem às informações que permitem que os dados possam ser armazenados, preservados e acessados, envolvendo três outras subcategorias que são os metadados técnicos, os metadados de direito e os metadados de preservação.

Em uma atualização, Riley (2017) acrescenta uma nova categoria às três categorias anteriormente mencionadas, denominada linguagem de marcação. O quadro 3 sintetiza essa divisão.

Quadro 3 - Tipos e funções dos Metadados

Tipo		Função
Metadados Descritivos		Encontrar ou entender um recurso
Metadados Administrativos	Metadados Técnicos	Decodificar e renderizar arquivos
	Metadados de Preservação	Gerenciar arquivos a longo prazo
	Metadados de Direitos	Garantir direitos de propriedade intelectual associados ao conteúdo
Metadados Estruturais		Descrever as relações das partes dos recursos
Linguagem de Marcação		Integrar metadados e sinalizar para outros recursos estruturais ou semânticos de conteúdo

Fonte: Adaptado de Riley (2017).

As comunidades de patrimônios culturais têm forte e antiga relação com os metadados descritivos, que estão relacionadas, principalmente, às informações intrínsecas dos recursos informacionais, possibilitando encontrá-los ou entendê-los. Já os metadados administrativos se referem às informações necessárias para gerenciar um recurso ou relacionadas à sua criação. Os metadados administrativos se subdividem em três. São eles: metadados técnicos- são as informações necessárias para decodificar e processar arquivos digitais; os metadados

de preservação, que consistem em informações para o gerenciamento a longo prazo, bem como para a migração futura ou emulação de arquivos digitais; e por fim metadados de direitos, que registram os direitos de propriedade intelectual associados aos recursos informacionais. Outro tipo são os metadados estruturais, que descrevem as relações entre as partes de um recurso. Por último, encontram-se as linguagens de marcação, que são linguagens atribuídas, que mesclam metadados ao conteúdo em forma de *tags*, as quais são inseridas para marcar características notáveis, sinalizando palavras com informações semânticas ou elementos estruturais (RILEY, 2017).

Apesar das várias interpretações sobre as funções que cada tipo de metadados desempenha, percebe-se, pela análise acima, que algumas se mantêm. Os metadados descritivos são uma categoria considerada unânime pela literatura analisada, sendo consensual sua importância e necessidade.

Outra categoria também recorrente é a dos metadados administrativos, que são comumente considerados uma grande categoria que abarca subcategorias e se destina a fornecer informações para o gerenciamento de um recurso digital ao longo de seu ciclo de vida. Nessa grande categoria é comum encontrar subtipos que se referem à preservação, às condições técnicas e aos direitos sobre a propriedade intelectual e o uso do recurso. Porém, autores como Pomerantz (2015), adicionam ainda categorias que relacionadas às informações sobre a proveniência dos dados e sobre os próprios metadados usados (*meta-metadada*).

Os metadados estruturais, em alguns casos, são considerados como parte da categoria de metadados administrativos. Contudo, em outros casos há uma categoria própria para eles. Em ambos, sua importância está em descrever as relações das partes dos recursos, de modo a definir sua interação com o todo ou com demais recursos.

No que diz respeito aos metadados de uso, há algumas divergências em sua categorização, contudo, sua finalidade está em registrar o nível e o tipo de uso dos recursos informacionais, o que é frequentemente considerado pela literatura.

Devido principalmente a sobreposição de funções que os metadados assumem, podem existir diferentes categorizações. No entanto, o importante é compreender que cada um desses tipos de metadados auxiliam nos diferentes casos de uso em sistemas de informação, seja para a descoberta, a identificação ou a compreensão de recursos informacionais, bem como a interoperabilidade entre os sistemas (RILEY, 2017). Nessa perspectiva, pode-se destacar a complexidade de processos a que os metadados estão associados, envolvendo seus tipos e funções.

Segundo Gilliland (2016, não paginado, tradução nossa) “a criação e o gerenciamento de metadados se tornaram uma mistura complexa de processos e camadas manuais e automáticos criados para muitas funções e indivíduos diferentes em diferentes pontos durante o ciclo de vida de um objeto de informações”. Por essa razão, além de seus diferentes tipos e funções, os metadados ainda apresentam características que os distingue quanto à fonte, ao método de criação, à natureza, à estrutura, ao *status*, à semântica e ao nível. Essas características são sistematizadas por Gilliland (2016).

Em relação às fontes de metadados, duas características são identificadas: os metadados podem ser intrínsecos ao recurso, ou seja, metadados que são gerados no momento da criação ou da digitalização de um recurso digital, por exemplo, o nome e o formato de um arquivo; e os metadados externos, gerados em um momento posterior à criação do recurso, de modo extrínseco, como os registros de localização, de descrição ou de direitos.

O método de criação dos metadados pode ser: automático, em que a criação, captura ou inferência ocorre por meio dos computadores; manual, gerados por seres humanos especialistas de informação; de forma manual ou automática durante processos de digitalização; ou ainda a partir da contribuição de usuários.

Em relação à natureza dos metadados, estes podem ser não especializados, criados por pessoas que não são especialistas ou profissionais da informação, por exemplo, usuários ou o próprio criador do recurso informacional; ou ainda podem ser especializados, criados por profissionais da informação e especialista no assunto ou domínio.

Referente à estrutura, os metadados podem ser apresentados de tal modo que se ajustam a uma estrutura padronizada ou proprietária previsível, como um formato de metadados ou de banco de dados específicos; ou podem não ser estruturados, não estando em uma estrutura previsível, tal como anotações de texto livre.

Para os *status* dos metadados, identificam-se aqueles que são estáticos, ou seja, que não são ou não devem ser alterados depois de criados; e aqueles que são dinâmicos, que podem mudar em razão do uso, manipulação ou para fins de preservação de um recurso; também podem ser categorizados em metadados de longo prazo, necessários para garantir que o recurso digital continue acessível e utilizável, por exemplo, informações de processamento e de preservação; metadados de curto prazo, principalmente de natureza transacional, por exemplo, informações de localização temporária; e, ainda, os metadados que são criados usando um sistema anterior de esquema de metadados, denominados metadados legados ou herdados.

Em relação à semântica, os metadados podem ser do tipo controlados, que se ajustam a um formulário padronizado de vocabulário ou de autoridade, seguindo regras de padrão de conteúdo; ou metadados não controlados, que não estão em conformidade com nenhum formulário padronizado de vocabulário ou de autoridade.

Por último, os metadados podem ser categorizados quanto ao nível, sendo diferenciados como metadados em nível de coleção ou grupo, que estão relacionados às coleções ou aos agrupamentos de recursos informacionais; e metadados em nível do item ou do item relacionado a itens individuais, geralmente contidos em coleções. Com base nas definições, o quadro 4 sintetiza as características que distingue os metadados quanto às categorias pontuadas por Gilliland (2016).

Quadro 4 - Características dos metadados

Categoria	Características
Fonte de metadados	Intrínsecos – metadados gerados no momento da criação ou digitalização de um recurso informacional.
	Extrínsecos - metadados gerados em um momento posterior à criação do recurso informacional.
Método de criação de metadados	Automático - metadados gerados por meio dos computadores.
	Manual - metadados gerados por seres humanos especialistas de informação.
	Manual ou Automático – metadados gerados durante processos de digitalização.
	Manual – gerados pela contribuição de usuários.
Natureza dos metadados	Não especializados – metadados criados por pessoas que não são especialistas ou profissionais da informação.
	Especializados - metadados criados por profissionais da informação ou especialistas no assunto ou domínio.
Estrutura	Estruturados – metadados que se ajustam a uma estrutura padronizada ou proprietária previsível.
	Não estruturados – metadados que não estão em conformidade com uma estrutura previsível.
<i>Status</i>	Estáticos - metadados que não são ou não devem ser alterados depois de criados
	Dinâmicos - metadados que podem mudar com o uso, manipulação ou para fins de preservação de um recurso.
	De longo prazo - metadados necessários para garantir que o recurso digital continue acessível e utilizável.
	De curto prazo – metadados de natureza transacional ou temporária.
	Herdados – metadados provenientes de um esquema de metadados anterior.
Semântica	Controlados - metadados que se ajustam a um formulário padronizado de vocabulário ou de autoridade e que seguem regras padrão de conteúdo.
	Não controlados - que não estão em conformidade com nenhum formulário padronizado de vocabulário ou de autoridade.
Nível	Coleção ou grupo - metadados que estão relacionados a coleções ou agrupamentos de recursos informacionais.
	Item individual - metadados que estão em nível do item ou do item relacionado a itens individuais, geralmente contidos em coleções.

Fonte: Adaptado de Gilliland (2016).

Tanto os tipos e as funções, quanto às características dos metadados revelam a complexidade de relações que os metadados estabelecem com os recursos informacionais durante seus ciclos de vida nos ambientes digitais (GILLILAND, 2016), envolvendo uma gama de atividades, como descrição, descoberta, gerenciamento, manutenção, preservação etc.

Devido a essa complexidade, um planejamento cuidadoso das estruturas que regem os metadados é recomendado. Boas estruturas de dados são criadas quando se segue um planejamento cuidadoso e princípios sólidos (ZENG; QIN, 2016). Tais princípios, segundo Duval *et al.* (2002, não paginado, tradução nossa), “[...] são aqueles conceitos julgados comuns a todos os domínios de metadados e que podem orientar o planejamento de qualquer esquema ou aplicativo de metadados”. Cada princípio refere-se a importantes requisitos a serem levados em consideração para o efetivo funcionamento dos sistemas informacionais digitais. Esses princípios correspondem, de acordo com Duval *et al.* (2002), à modularidade, à extensibilidade, ao refinamento e ao multilinguismo e estão intimamente ligados à interoperabilidade.

O princípio da modularidade consiste na capacidade de que elementos de dados de diferentes padrões, bem como vocabulários e outros metadados, possam ser combinados de maneira sintática e semanticamente interoperável. Esse princípio permite que novas estruturas sejam criadas com base em padrões já estabelecidos, tirando proveito das melhores práticas já observadas. Na arquitetura de metadados da *Web*, esse princípio deve permitir certa flexibilidade, possibilitando combinar uma variedade de módulos semânticos dentro de uma base sintática comum (DUVAL *et al.*, 2002; ZENG; QIN, 2016). Isso significa que é possível reutilizar os metadados existentes, uma vez que estes se apresentam em módulos que podem ser combinados e recombinaos em outras estruturas. Tais módulos permitem uma abordagem baseada em perfis de aplicativos, que vão propiciar que elementos de dados extraídos de uma ou mais estruturas sejam adequados para uma aplicação específica na *Web*.

O princípio da extensibilidade consiste na capacidade de permitir estender o número de metadados de acordo com as necessidades, possibilitando “[...] adaptar uma determinada aplicação às necessidades locais ou às necessidades específicas do domínio sem comprometer indevidamente a interoperabilidade fornecida pelo esquema base” (DUVAL *et al.*, 2002, não paginado, tradução nossa). Ou seja, consiste em permitir a inclusão de novos elementos sem, entretanto, comprometer os elementos já existentes.

O princípio de refinamento refere-se ao grau de detalhamento a ser considerado mediante o domínio de aplicação, permitindo escolher o nível de descrição apropriado e

suficiente para cada situação. Para esse princípio, duas noções devem ser consideradas: uma refere-se à inclusão de qualificadores para refinar ou especificar o significado de um elemento de metadados; e a outra, refere-se à especificação de padrões de codificação de valores a serem usados para um determinado elemento de metadados (DUVAL *et al.*, 2002; ZENG; QIN, 2016).

Por fim, o multilinguismo indica a capacidade de incluir os aspectos da diversidade linguística e cultural na padronização dos metadados. É preciso garantir tanto a utilização de padrões que sejam internacionais, quanto de padrões que sejam adaptados a contextos e convenções locais. Na *Web*, os padrões devem suportar ambas as estratégias, bem como a combinação das duas, viabilizando um equilíbrio para um esquema global (DUVAL *et al.*, 2002). O quadro 5 sintetiza cada um desses princípios.

Quadro 5 - Princípios dos metadados

Modularidade	Capacidade dos metadados serem combinados de maneira sintática e semanticamente interoperável de modo que novas estruturas possam ser criadas com base em padrões já estabelecidos.
Extensibilidade	Capacidade de estender o número de metadados de acordo com as necessidades locais específicas do domínio, permitindo a inclusão de novos elementos sem comprometer os elementos já existentes.
Refinamento	Capacidade de escolher o nível de descrição apropriado e suficiente para cada domínio de aplicação, levando em consideração a inclusão de qualificadores e a escolha de padrões de valores.
Multilinguismo	Capacidade de incluir os aspectos da diversidade linguística e cultural na padronização dos metadados, contemplando necessidades gerais e locais por meio de um esquema global.

Fonte: Elaborado pela autora com base em Duval *et al.* (2002).

Diante desses princípios, reafirma-se a importância do planejamento e da estruturação padronizada dos metadados para que os dados sejam úteis e válidos em um sistema de informação. Zeng e Qin (2016, p. 28, tradução nossa) destacam que tais princípios visam que as implementações de metadados sejam “[...] tanto sustentáveis para a utilização e preservação a longo prazo, quanto interoperáveis para compartilhamento e reutilização”.

Nessa perspectiva, destaca-se que tanto a sustentabilidade, quanto a interoperabilidade têm sido consideradas questões-chave para os sistemas de informação que lidam com grandes volumes de dados heterogêneos e distribuídos.

A sustentabilidade pode ser entendida, no contexto da preservação digital, como “[...] a habilidade de identificar os recursos necessários para manter o valor de um objeto digital ou

de um serviço para os usuários aos quais se destinam” (MARON; YUN; PICKLE, 2013, p. 12, tradução nossa). Como explicam Maron, Yun e Pickle (2013), muitos conteúdos podem estar sob ações que os mantêm seguros e íntegros em servidores locais, porém sem um suporte contínuo visando o interesse de seu público-alvo; esses conteúdos podem ter seu impacto reduzido tornando-se invisíveis para um buscador ou, então, incompatíveis com a tecnologia atual e com os dispositivos e serviços. Por isso, a sustentabilidade está relacionada com as ações que mantêm o valor e a relevância de um recurso para uso futuro.

Segundo Zeng e Qin (2016, p.36, tradução nossa):

Uma estratégia fundamental para o acesso sustentável e efetivo às informações é ter boa qualidade de metadados vinculáveis prontos para suportar as necessidades de gerenciamento, recuperação, navegação, descoberta, uso e reutilização de recursos em qualquer dispositivo, de maneira responsiva e dinâmica.

Pode-se dizer que o mesmo vale em relação à interoperabilidade. A interoperabilidade é entendida como a capacidade de vários sistemas trocarem dados com a mínima perda de conteúdo e funcionalidade (RILEY, 2016). Para isso, é necessário ter metadados de boa qualidade que possam suportar os processos de mapeamento e busca de dados em diferentes sistemas.

Para alcançar a interoperabilidade, o principal desafio é que os problemas derivados da heterogeneidade dos dados precisam ser eliminados. Os dados podem ser heterogêneos em sintaxe, esquema ou semântica (CRUZ; XIAO, 2005). De acordo com Cruz e Xiao (2005, p. 2, tradução nossa), “A heterogeneidade sintática é causada pelo uso de diferentes modelos ou linguagens. A heterogeneidade esquemática resulta de diferenças estruturais. A heterogeneidade semântica é causada por diferentes significados ou interpretações de dados em vários contextos”.

Ao longo dos anos, diversos tipos de padrões foram desenvolvidos com a finalidade de atender necessidades de representação da informação e do conhecimento de domínios específicos. Porém, essa diversidade de padrões existentes gera uma variedade de formatos para os dados, resultando em diferentes estruturas e terminologias conflitantes (CASTRO; SIMIONATO; ZAFALON, 2016), o que dificulta a interoperabilidade.

Como destaca Hyvönen (2012, p.42, tradução nossa), “[...] os problemas de interoperabilidade podem ser resolvidos efetivamente usando um único esquema. No entanto, esquemas diferentes são necessários e usados para diferentes tipos de dados em sistemas que lidam com conteúdo entre domínios.” Assim, a interoperabilidade se torna um posto-chave

para que os sistemas que suportam a integração e o intercâmbio de dados, tal como no ambiente *Web*.

Principalmente, se considerar os dados de coleções de patrimônios culturais, que têm como uma das suas principais características a heterogeneidade sintática e semântica que provem dos vários tipos de relações e eventos que envolvem esses dados.

Nesse contexto, a interoperabilidade está relacionada à capacidade de maximizar o valor e o potencial de reuso da informação e de intercambiar efetivamente tais informações de modo que novos conhecimentos possam ser gerados a partir dos diversos relacionamentos entre os dados. Ainda, a interoperabilidade permite que recursos informacionais heterogêneos, armazenados em diferentes bases de dados, possam ser buscados e recuperados utilizando uma única interface (SAYÃO; MARCONDES, 2008).

Como esclarece Hyvönen (2012, p. 42, tradução nossa, grifo nosso):

A interoperabilidade sintática pode ser obtida pela harmonização de formulários estruturais para representar dados (por exemplo, usando elementos DC [**Dublin Core**] ou por um esquema XML) e fixando convenções de codificação de valores para elementos (por exemplo, vocabulários controlados, formato de datas, sistema de coordenadas etc.). A interoperabilidade semântica é obtida por convenções compartilhadas para interpretar as representações sintáticas, por exemplo, que a propriedade *dc:subject* descreve o assunto de um documento como um conjunto de recursos de palavras-chave retirados de uma ontologia.

No ambiente *Web*, esses fatores têm sido resolvidos a partir dos padrões recomendados pelas agências internacionalmente reconhecidas, como a W3C (*World Wide Web Consortium*) e o DCMI (*Dublin Core Metadata Initiative*), que têm como objetivo viabilizar e promover a representação e o compartilhamento de dados estruturados.

A recomendação de formatos em linguagem XML (*eXtensible Markup Language*) tem permitindo resolver as questões relacionadas à padronização sintática para a codificação dos dados (DOERR, 2003; CRUZ; XIAO, 2005), apesar de existirem outras linguagens mais simples com a mesma função da XML. Em relação à interpretação semântica, as soluções envolvem os modelos de ontologia conceitual, construídos com base nas tecnologias recomendadas para a *Web* e nos princípios semânticos de cada domínio (DOERR, 2003; HYVÖNEN, 2012).

Por esse ponto de vista, vislumbra-se o importante papel dos metadados em todos os aspectos da representação e da organização da informação, em especial nos ambientes digitais na *Web*, orientados para a construção semântica, a interoperabilidade e a integração dos dados.

2.2 Os metadados e as tecnologias semânticas

Uma das importantes características dos dados em ambientes digitais abertos como na *Web*, é o fato de que eles nem sempre estão estruturados de forma coerente para serem utilizados e disponibilizados aos usuários. Porém, isso não significa que o conteúdo publicado na *Web* seja gerado de modo desestruturado.

Segundo Antoniou *et al.* (2012, p. 2, tradução nossa):

A grande maioria do conteúdo da web está sendo gerada a partir de bancos de dados e sistemas de gerenciamento de conteúdo contendo conjuntos de dados cuidadosamente estruturados. No entanto, a estrutura muitas vezes rica que está disponível nesses conjuntos de dados é quase completamente perdida no processo de publicação de dados estruturados como páginas HTML (*Hypertext Markup Language*) legíveis por humanos.

Tal situação é particularmente notável no universo das bibliotecas, dos arquivos e dos museus, que tradicionalmente criam estruturas cuidadosamente padronizadas para representar seus dados, mas enfrentam desafios ao buscarem o ambiente *Web* como meio de disponibilizá-los.

Esses desafios são decorrentes principalmente das estruturas variáveis de dados. Ao serem inseridos na *Web*, esses dados precisam atender a estruturas que explicitem seus aspectos semânticos de modo compatível com o ambiente informacional. Para que isso seja possível, é necessário considerar as particularidades e diferenças entre as estruturas e níveis de especificidade, bem como a finalidade para a qual uma estrutura é desenvolvida (ALVES, 2010). No caso da *Web*, estruturas mais ricas, ou seja, altamente estruturadas e constituídas por metadados complexos e específicos, acabam por não ser totalmente legíveis por seus sistemas, gerando problemas de interoperabilidade.

Sendo esses desafios recorrentes a demais campos de atuação, as questões relacionadas a estruturação dos dados na *Web* tornou-se uma das grandes iniciativas do grupo de Tim Berners-Lee⁵, tendo como objetivo que os dados fossem bem definidos e ligados de forma legível para processamento em máquinas, visando à automação, à integração e à reutilização dos dados (BERNERS-LEE, 2001). O empreendimento popularizou-se com o nome de *Web Semântica* e a ideia fundamental por trás do termo é a de que uma *Web* baseada em um maior nível de estruturação semântica permitiria publicar e interligar os conjuntos de dados conjuntamente com as estruturas subjacentes (ANTONIOU *et al.*, 2012).

⁵ Berners-Lee é o criador e diretor do *World Wide Web Consortium* (W3C), e seu grupo supervisiona o desenvolvimento contínuo da *Web*.

Como a denominação sugere, a *Web Semântica* tem como intuito introduzir a estruturação do conteúdo dos recursos digitais por meio da definição semântica dos dados. Diferente da clássica *Web* de documentos que tem por base as linguagens de marcação para a apresentação do seu conteúdo em formato *Hypertext Markup Language* (HTML) (MORI; CARVALHO, 2004), a *Web Semântica* visa, assim, trazer estrutura ao conteúdo disponibilizado na *Web* por meio de ferramentas padronizadas de representação semântica de dados.

Berners-Lee, Hendler e Lassila (2001) explicam que a *Web Semântica* é uma extensão na arquitetura da *Web*, na qual a informação nela contida passa a ter significado bem definido e compreendido pelas máquinas. Essa extensão tem sido constantemente aprimorada pelos desenvolvedores da *World Wide Web Consortium* (W3C) para sustentar a chamada *Web de Dados* e introduzir maior interação semântica entre homens e máquinas.

Segundo Antoniou *et al.* (2012, p. 2, tradução nossa) a *Web Semântica* segue princípios que podem ser resumidos da seguinte forma:

1. disponibilizar dados estruturados e semi-estruturados em formatos padronizados na web;
2. tornar não apenas os conjuntos de dados, mas também os elementos de dados individuais e suas relações acessíveis na web;
3. descrever a semântica pretendida de tais dados em um formalismo, para que essa semântica pretendida possa ser processada por máquinas.

Tais princípios se sustentam por meio das chamadas tecnologias semânticas, que viabilizam a representação e o compartilhamento de dados de modo estruturado, bem definido e principalmente padronizado. Segundo Pomerantz (2015), não só a estruturação de dados é necessária para alcançar a visão da *Web Semântica*, mas também que essa estruturação seja aderente a padrões amplamente compartilhados.

Na mesma linha de pensamento, Antoniou *et al.* (2012, p. 4, tradução nossa, grifo do autor) faz um paralelo entre os três princípios que caracterizam a *Web Semântica*, mencionados acima, e as soluções até então desenvolvidas, de modo correspondente.

1. usar grafos rotulados como o modelo de dados para objetos e suas relações, com os objetos como os nós do grafo e as arestas representando as relações entre esses objetos. [...] “Resource Description Framework” RDF é usado como o formalismo para representar tais grafos.
2. usar identificadores da web (Uniform Resource Identifiers - URI) para identificar os itens de dados individuais e suas relações que aparecem nos conjuntos de dados. Novamente, isso é refletido no design do RDF.
3. usar ontologias (resumidamente: vocabulários hierárquicos de tipos e relações) como o modelo de dados para representar formalmente a semântica pretendida dos dados. Formalismos como RDF Schema e The Web Ontology

Language (OWL) são usados para essa finalidade, novamente usando URIs para representar os tipos e suas propriedades.

Tais soluções se destacam por proporcionarem a padronização nas estruturas de dados e por permitirem estabelecer as ligações de significado entre eles de tal forma que as máquinas tenham maior capacidade de processar e trocar informações semanticamente.

De acordo com Santarem Segundo (2015, p. 223-224),

Desde a proposta inicial de 2001, a *Web Semântica* vem ganhando força e agregando novas tecnologias, funcionalidades e evoluindo para tornar real o processo de construção de ambientes semânticos. Tecnologias como RDF (*Resource Description Framework*), XML (*eXtensible Markup Language*), OWL (*Web Ontology Language*) e todos os conceitos que as envolvem, ganham novas versões, descritos com clareza no W3C (*World Wide Web Consortium*), e tornam possível a materialização do conceito da *Web Semântica*.

De modo dinâmico, a *Web Semântica* tem sido aprimorada a cada dia para dar suporte a um conjunto de tecnologias orientadas à padronização da representação semântica dos dados disponíveis na *Web*. Uma das importantes ferramentas utilizadas não só para a proposta da *Web Semântica*, mas para *Web* como um todo, são as linguagens de codificação, que dizem respeito à estrutura sintática da representação das informações, tendo como uma das recomendações a linguagem *Extensible Markup Language* (XML).

De acordo com Coneglian *et al.* (2019, não paginado),

[...] a linguagem XML pode ser compreendida como a sintaxe utilizada para a disponibilização das informações, que permite agregar semântica aos documentos e possibilita que cada aplicação crie uma interpretação da marcação atribuída ao conteúdo.

Devido a sua funcionalidade flexível, aberta e independente de dispositivo, a XML apresenta grande capacidade de interoperar (SANTAREM SEGUNDO, 2004) e por isso tem sido muito utilizada no decorrer dos anos. Porém, outras linguagens também vêm sendo utilizadas, por exemplo, o formato *Turtle* (*Terse RDF Triple Language*) que é uma sintaxe para expressar dados no modelo de dados do RDF (WORLD WIDE WEB CONSORTIUM, 2011) ou o formato JSON (*JavaScript Object Notation*) que consiste em uma formato leve para troca de dados baseada em um subconjunto da linguagem de programação *JavaScript* (JSON, [2020]). Linguagens como a XML, *Turtle* e JSON permitem definir a estrutura das informações nas páginas da *Web* de tal modo a criar uma representação mais facilmente processável por máquinas.

Além disso, outra tecnologia base da *Web Semântica* é o *Resource Description Framework* (RDF), uma arquitetura de metadados que permite fazer declarações sobre os

dados e seus relacionamentos. Configurando-se como um modelo de dados simples, o RDF destina-se a descrever semanticamente recursos na *Web* possibilitando fazer declarações sobre eles.

O formato dessas declarações é constituído de três elementos: sujeito, predicado e objeto, denominado triplas. Essas triplas também são denominadas como ‘recurso, propriedade e valor’ ou ainda, ‘entidade, atributo, valor’, formando os chamados grafos (BIZER; HEATH; BERNERS-LEE, 2009). A função das triplas é expressar um relacionamento entre dois recursos, sendo que um deles é o sujeito e o outro o objeto e entre ele está o predicado, também chamado de propriedade, que representa a natureza do relacionamento entre os recursos (WORLD WIDE WEB CONSORTIUM, 2014).

Tanto os recursos, quanto as propriedades, são descritos usando um *Uniform Resource Identifier* (URI). Um URI consiste em um identificador uniforme que permite identificar um recurso digital na *Web*, de tal modo que esses recursos possam ser associados a partir de uma rede de relacionamentos (WORLD WIDE WEB CONSORTIUM, 2006). Os URIs são a base para o estabelecimento das relações entre os recursos na *Web Semântica*, pois referenciam de modo unívoco os recursos e suas propriedades, auxiliando na descrição e identificação das informações.

Uma das principais tecnologias para apoiar a construção de uma rede de relacionamentos semânticos, são as ontologias, que consistem em modelos formais relacionados ao modo como se percebe os possíveis estados de coisas em um determinado domínio do discurso (GRUBER, 1994, GUARINO, 1998 *apud* DOERR, 2008). Elas podem ser entendidas como um modelo de relacionamento de entidades e suas interações que se referem a algum domínio particular do conhecimento (SOUZA; ALVARENGA, 2004).

Segundo Castro (2012), para campo da Ciência da Informação, o interesse nos estudos sobre ontologias se fundamenta:

[...] na potencialidade que elas tem em organizar e representar a informação, principalmente no que diz respeito ao melhoramento dos processos de recuperação da informação em múltiplos domínios e sua aplicação efetiva nos variados ambientes informacionais digitais.

Para tanto, as ontologias são uma forma de construir “[...] uma relação organizada entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando mais eficiente e facilitando o processo de interpretação dos dados pelas ferramentas de recuperação da informação” (SANTAREM SEGUNDO, 2015, p. 226).

O potencial das ontologias pode ser percebido no fato de que elas configuram-se como instrumentos de explicitação do processo de representação de conceitos e de seus relacionamentos dentro de um domínio (RAMALHO, 2006).

Os modelos conceituais possibilitam, por sua vez, descrever a estrutura de um conjunto de dados, incluindo os dados, as características dos dados, bem como os relacionamentos e as regras para esses dados (FUSCO, 2010). Assim, a modelagem conceitual “[...] tem como objetivo identificar, modelar e implementar um modelo de dados consistente com as necessidades do domínio expressas na especificação de requisitos (FUSCO, 2010, p. 95).

Para isso, é necessário que as ontologias passem de uma estrutura conceitual para uma linguagem de implementação (SANTAREM SEGUNDO; CONEGLIAN, 2016). Existem uma variedade de linguagens já desenvolvidas com essa finalidade, porém a recomendação mais recente da W3C é a *Ontology Web Language* (OWL).

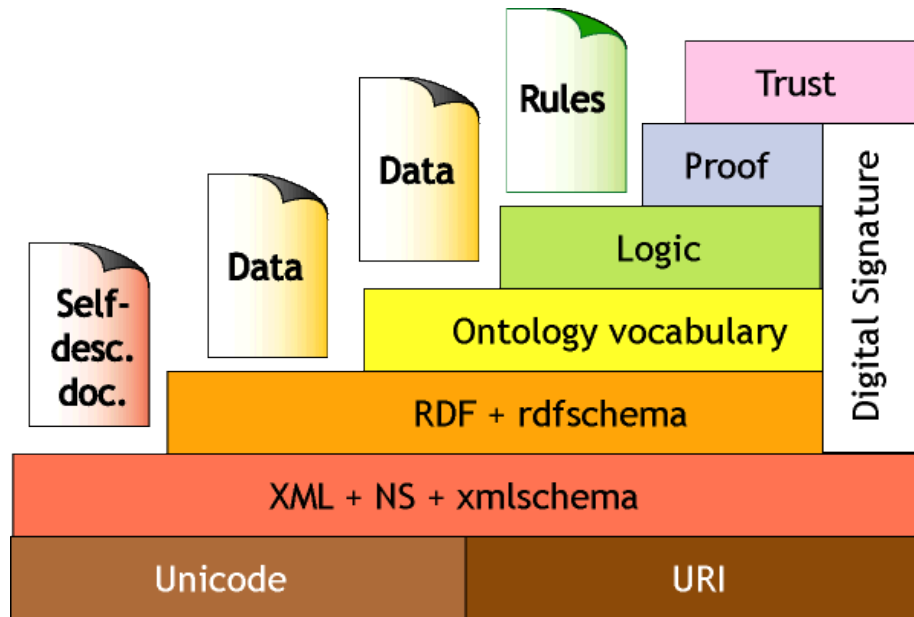
A OWL configura-se como “[...] uma linguagem de marcação semântica para a definição, a instanciação, a publicação e a partilha de ontologias na *World Wide Web*” (SANTAREM SEGUNDO; CONEGLIAN, 2016, p. 221). Isso permite que as informações e as especificações, ou seja, que as representações de um domínio possam ser interpretadas de modo compartilhado a partir da modelagem dos aspectos semânticos desse domínio.

As tecnologias presentes na arquitetura da *Web Semântica* trabalham de modo integrado para que os recursos sejam bem representados semanticamente pelos metadados e pelas ontologias, permitindo que as ferramentas de busca tenham maior capacidade de recuperar as informações de modo eficiente.

Os metadados têm a função de descrever os recursos informacionais disponibilizados na *Web* e, as ontologias, de estruturar os aspectos semânticos das informações destes recursos (SANTOS; ALVES, 2009). Isso implica que, na *Web*, as ontologias configuram-se como a base semântica para os metadados, de modo a explicitar para as máquinas os elementos que caracterizam um recurso informacional.

A *Web Semântica* é composta por diversas tecnologias que trabalham em diferentes aspectos necessários à representação da informação. A forma mais comum de apresentar a arquitetura dessas tecnologias é o chamado “bolo de noiva”, que consiste em uma estrutura gráfica disposta em camadas sobrepostas e escalonáveis que descrevem linguagens e tecnologias que são necessárias à proposta *Web Semântica*, como apresentado na figura 1.

Figura 1 - “Bolo de noiva” da *Web Semântica*

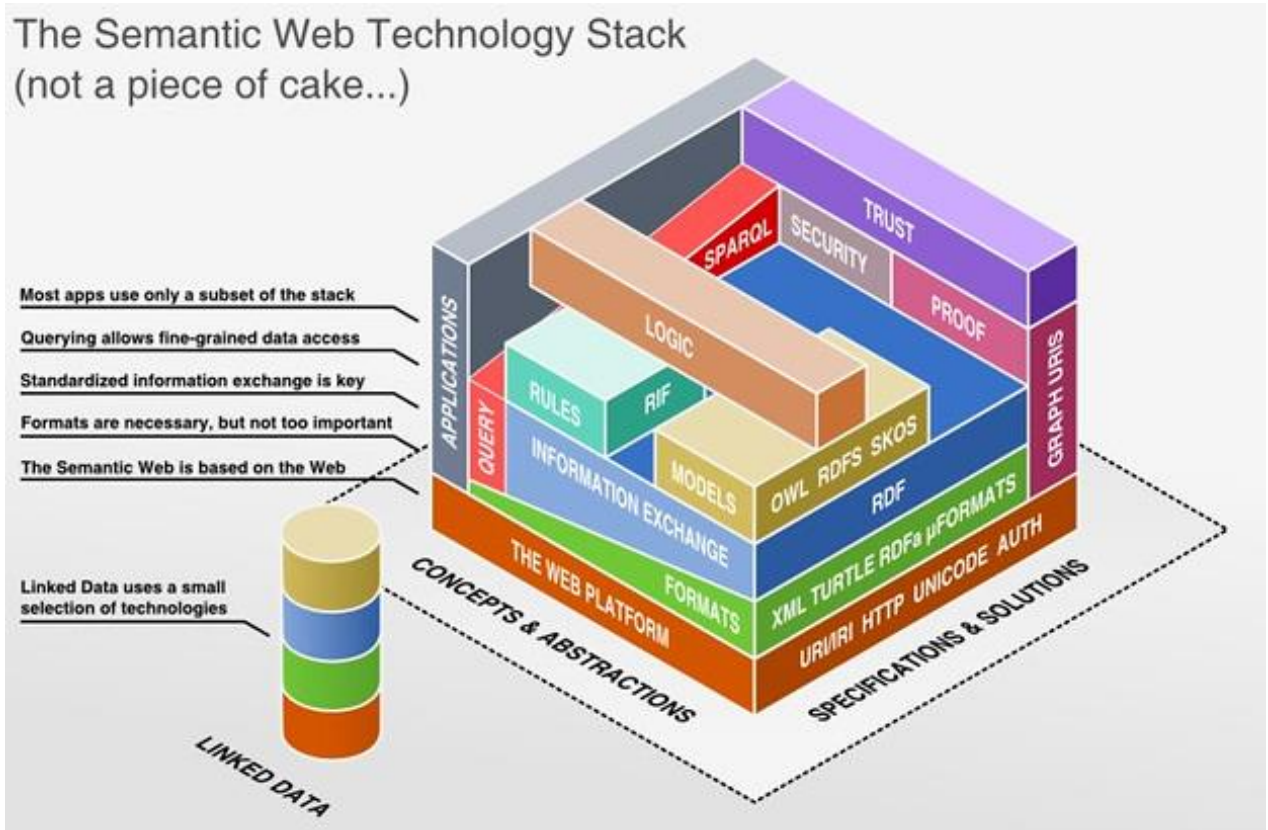


Fonte: Berners-Lee (2000).

Desde a proposta inicial da *Web Semântica*, muitas mudanças ocorreram em relação as tecnologias recomendadas, refletindo na sua arquitetura proposta. Após sucessivas atualizações nessa arquitetura, tem-se, mais recentemente, o diagrama das camadas proposto por Nowack (2009), apresentado na figura 1, no qual é possível identificar as tecnologias e as recomendações já estabelecidas até o momento.

Nessa abordagem, Nowack (2009) apresentou uma nova forma de estrutura, por meio de um diagrama no formato de cubo, no qual de um lado é possível ver os conceitos e as abstrações e de outro as especificações e as soluções tecnológicas desenvolvidas para cada necessidade. Com isso, o autor enfatiza que a arquitetura da *Web Semântica* não deve ser representada como um “uma fatia de bolo”, mas como uma “pilha de tecnologias”, demonstrando a complexidade das tecnologias envolvidas em sua proposta.

Figura 2 - Tecnologias da Web Semântica



Fonte: Nowack (2009).

Na figura 2, as tecnologias semânticas podem ser visualizadas em cada uma das camadas, a partir de suas especificações e soluções. Na camada base encontram-se as soluções já conhecidas e estabelecidas para a plataforma da *Web* (*The web platform*), que tem como solução o URI/IRI, HTTP, UNICODE e AUTH, responsáveis pela identificação e codificação dos recursos na *Web*.

Na camada seguinte, encontram-se os formatos (*Formats*) que contemplam a estrutura da representação das informações, tendo como recomendação a *Extensible Markup Language* (XML), porém havendo a possibilidade de utilizar outros formatos como *Turtle* e *Resource Description Framework in Attributes* (RDFa) (NOWACK, 2009).

A próxima camada é estabelecida pelo *Resource Description Framework* (RDF), uma arquitetura de metadados que tem a finalidade de auxiliar a troca de informações (*Information exchange*), sendo considerada por Nowack (2009) como a chave para a proposta. Trabalhando de modo integrado com as outras tecnologias, o RDF viabiliza que a interface de busca (*Query*) acesse os dados por meio do *Protocol and RDF Query Language* (SPARQL), ferramenta que recupera dados em RDF.

A camada denominada modelos (*Models*), está relacionada às ontologias e aos modelos de dados, e tem a finalidade de auxiliar na representação dos aspectos semânticos e dos relacionamentos entre os dados, bem como permitir, a partir desses dados, a realização de inferências. Nessa camada, as soluções envolvem o *Web Ontology Language* (OWL), o *Resource Description Framework Schema* (RDFS) e o *Simple Knowledge Organization System* (SKOS). Alinhado a essa camada estão as regras (*Rules*), definidas pelo *Rule Interchange Format* (RIF), que estabelecem “línguas de regras” (WORLD WIDE WEB CONSORTIUM, 2013a), necessárias para seu intercâmbio no sistema. Também relacionado à essa camada, estão as interfaces de segurança (*Security*) e de prova (*Proof*), representadas pelos denominados GRAPH URIs, que são conjuntos de delações em RDF identificados com URIs (Identificador Uniforme de Recursos), que permitem a descrição de um conjunto informações, como contexto, proveniência, entre outros metadados.

Por fim, encontra-se a camada de confiança (*Trust*) que é formada com base nas camadas anteriores, permitindo verificar o grau de confiabilidade das declarações feitas a partir das informações representadas.

Diante do exposto, precebe-se que a proposta da *Web Semântica* é um empreendimento com desenvolvimento gradativo, que vem se expandindo na medida em que as tecnologias vão permitindo dar maior nível semântico aos dados disponibilizados na *Web*.

Tais tecnologias, por sua vez, tem por base os metadados. Entendidos no contexto da *Web* como dados processáveis ou interpretáveis por máquinas (HYVÖNEN, 2012), os metadados são essenciais para a proposta da *Web Semântica*, sobretudo porque seu propósito é além de dar acesso aos dados, é também estabelecer os relacionamentos entre eles. Esses relacionamentos entre os dados ou conjuntos de dados são o cerne da *Web Semântica*, possibilitando a reutilização dos dados já publicados e constituindo o denominado *Linked Data*.

Apresentado no diagrama de Nowack (2009), (Figura 2), o *Linked Data* refere-se a um conjunto de práticas, que foi desenvolvido como parte da proposta da *Web Semântica* de publicar e conectar dados estruturados na *Web* (CUNHA; SOUZA; LÓSCIO, 2011, p.79). O *Linked Data* baseia-se em tecnologias como URI para a identificação de recursos e no RDF para fornecer informações sobre os recursos informacionais e seus relacionamentos, conectando-os a outros recursos na *Web*.

Baker (2012, p. 117, tradução nossa) corrobora, com base Berners-Lee (2009), que o

[...] *Linked Data* é um conglomerado crescente de conjuntos de dados na rede mundial que está interligado por meio do *Description Resource*

Framework (RDF) usando identificadores de recursos uniformes baseados na *Web* (URIs, ou ‘web endereços’) para designar as coisas descritas e os termos usados para descrevê-los.

Alinhado a proposta do *Linked Data*, encontra-se o *Linked Open Data*, que se refere aos dados que são vinculados sob uma licença aberta. De acordo com Baker *et al.* (2011), o *Linked Data* diz respeito à interoperabilidade técnica dos dados, enquanto que *Open Data* centra-se na sua interoperabilidade legal. Ou seja, “[...] a tecnologia *Linked Data*, por si só, não exige que os dados sejam abertos, embora o potencial da tecnologia seja melhor percebido quando os dados são publicados como *Linked Open Data*.” (BAKER *et al.*, 2011, não paginado, tradução nossa). Ressalta-se assim que os princípios *Linked Data* e *Open Data* são ideologias adotadas para melhorar a conectividade e a recuperação dos dados e informações, sendo que a fusão desses princípios origina o movimento *Linked Open Data* (LOD).

Harpring (2016, p.35) explica que:

Ter dados interligados e abertos significa que eles são estruturados e publicados como descrito pelo *World Wide Web Consortium* (W3C). Isso permite que os dados sejam interligados e disponibilizados abertamente, além de compartilháveis na *Web Semântica*. O objetivo dos LOD é tornar os dados mais úteis, permitindo que diferentes recursos sejam interconectados e pesquisados.

Desse modo, para que a *Web Semântica* esteja mais próxima da realidade, é necessário “[...] que se tenha uma grande quantidade de dados disponíveis na *Web* em formatos padronizados, recuperáveis e manuseáveis pelas ferramentas da *Web Semântica*” (WORLD WIDE WEB CONSORTIUM, 2015, não paginado, tradução nossa), o que, como pode ser percebido, reflete diretamente na utilização de metadados.

Sob essa perspectiva, é possível elucidar a relação existente entre as tecnologias semânticas e os metadados, uma vez que para atender aos requisitos de interoperabilidade, tanto funcional, quanto semântica, entre padrões de metadados heterogêneos, é necessário uma base conceitual que suporte e que defina os variados tipos de dados provenientes da representação do recurso informacional.

Discutidos os fundamentos que regem os metadados e sua implicação no ambiente *Web*, a seção seguinte contempla a revisão de literatura acerca da curadoria digital.

3 CURADORIA DIGITAL

Os ambientes informacionais digitais permitiram novas perspectivas e possibilidade de acesso e compartilhamento da informação, mudando as formas de atuação de diversos setores, principalmente, os relacionados com a produção e a disseminação de informação e de conhecimento. Aliados ao contínuo desenvolvimento das tecnologias, esses ambientes se transformaram em importantes meios pelos quais a informação circula, é compartilhada e recuperada. No entanto, a riqueza e a quantidade de recursos informacionais que passaram a ser acumulados em formatos digitais, tornam o tratamento informacional um desafio ainda maior, exigindo que as abordagens tradicionais fossem repensadas.

Como explica Abbott (2008, não paginado, tradução nossa), a existência de “[...] uma quantidade cada vez maior de dados criados em formatos digitais, por meio da digitalização de informações analógicas existentes e da criação de novos dados ‘natos digitais’ dos setores de ciências, artes e humanidades.”, fez crescer a preocupação em relação à obsolescência tecnológica e fragilidade inerente ao meio digital.

Em razão disso, surgiram os primeiros esforços para garantir a preservação digital. No entanto, a preservação, visando somente à integridade e autenticidade dos dados, implicava em pouco acesso. Com o tempo o foco mudou para que os recursos digitais pudessem ser preservados sem comprometer o seu acesso, passando então a serem gerenciados ao longo de todo o seu ciclo de vida (HIGGINS, 2011).

Abordagens em relação ao tratamento do material digital partiram inicialmente das comunidades de prática e pesquisa em preservação digital (KIM, 2014; DALLAS, 2016). Com o tempo, a onipresença dos dados passou a ser preocupação de mais comunidades, que viam o conhecimento humano ser cada vez mais produzido e mantido em formatos digitais, culminando no diálogo entre diferentes profissionais para o desenvolvimento de uma abordagem orientada aos dados.

Desse diálogo interdisciplinar, nasce um conjunto de estratégias, de práticas e de ferramentas voltada para a contínua gestão dos dados, com o foco além da preservação digital. Após várias discussões, tal abordagem acabou sendo denominada como curadoria digital. O termo curadoria digital foi usado pela primeira vez no “*Digital Curation: digital archives, libraries and e-science seminar*”, um evento internacional realizado em Londres, em outubro de 2001, para discutir questões relativas ao desenvolvimento no campo da curadoria de dados e preservação digital, que ao mesmo tempo oportunizou o diálogo entre arquivistas,

bibliotecários e especialistas em gestão de informação e gestores de dados em *e-science* (BEAGRIE; POTHEN, 2001, BEAGRIE, 2006).

Como destaca Beagrie (2006), o novo termo foi cuidadosamente escolhido com a intenção de transmitir a necessidade de uma nova abordagem para criar e gerenciar recursos digitais que, ao mesmo tempo, pudesse incorporar aspectos de conceitos já existentes e utilizados pelas comunidades científicas e de bibliotecas digitais, como “curadoria de dados” e “preservação digital”, respectivamente. Yakel (2007) aponta que embora os termos ‘curadoria de dados’ e ‘preservação digital’ tenham sido usados anteriormente, o interesse intenso em curadoria digital demonstrava uma necessidade emergente.

Santos (2014, p. 29) corrobora que,

Os conceitos de "curadoria" digital começaram a aparecer depois que a preservação digital e os seus desafios e limitações eram conhecidas e a Internet como meio global para comunicação e divulgação da informação (inclusive a científica) já estava consolidada. A curadoria digital evoluiu das noções de preservação digital e da necessidade da informação ser divulgada em meio aberto, principalmente a informação científica produzida em rede e de forma distribuída.

Logo, é possível perceber que junto ao movimento de disponibilização de informação que se intensificou com as possibilidades de publicação de dados em ambientes digitais, veio à tona a necessidade de se pensar nas formas de tratamento dessas informações em ambientes diversificados, dinâmicos e com estruturas variáveis. É nessa direção que se desenvolve as práticas de curadoria digital, atreladas às necessidades de se disponibilizar recursos informacionais em novos ambientes, de modo que a informação pudesse ser acessada de forma contínua, abrangente e distribuída. Com isso, o termo passou a ser cada vez mais usado para designar um conjunto de habilidades e competências voltadas ao gerenciamento contínuo dos dados.

Nas comunidades científicas e acadêmicas, o grande interesse em estratégias voltadas ao gerenciamento a longo prazo dos dados contribuiu significativamente para o desenvolvimento da curadoria digital. Como explica Harvey (2010, p.15, tradução nossa), muito dos avanços teóricos, práticas, ferramentas vem das comunidades científicas, acadêmicas e de pesquisa, que “[...] dependem cada vez mais da computação em rede para conectar pesquisadores e acadêmicos em todo o mundo e gerar e compartilhar grandes conjuntos de dados [...]”.

Por essa razão, nas últimas décadas, verifica-se a existência de uma variedade de estudos voltados para a curadoria digital de dados científicos e aplicados à pesquisas em *e-*

Science, como, por exemplo, os trabalhos de LORD *et al.*, 2004, BALL, 2010, SALES; SAYÃO, 2012, KIM, 2014, POOLE, 2016, WEBER, 2016, SAYÃO; SALES, 2012, 2016.

Contudo, como salienta Yakel (2007), a curadoria digital se mostrou uma área ativa, abarcando diversas comunidades, tanto científicas, educacionais e profissionais, como também organizações governamentais e privadas, sendo mencionada em relatórios de infraestrutura cibernética, apresentações de pesquisas em conferências importantes e em programas educacionais.

Nesse cenário, um importante fator que viria a ajudar a consolidar o termo curadoria digital e popularizar suas práticas nas diversas áreas de atuação foi a criação de um centro especializado em âmbito nacional no Reino Unido. O *Digital Curation Centre* (DCC) foi fundado em 2004, com o intuito de “[...] apoiar instituições britânicas que armazenam, gerenciam e preservam informações digitais, e para ajudá-las a abordar o aprimoramento e a disponibilidade contínua de informações digitais para uso a longo prazo.” (PENNOCK, 2006, p. 1, tradução nossa). O centro é um consórcio entre quatro parceiros principais: a Universidade de Edimburgo; o centro especializado UKOLN na Universidade de Bath; o Instituto de Tecnologia e Informação Avançada em Humanidades (HATII) da Universidade de Glasgow, e; o Conselho do Laboratório Central dos Conselhos de Pesquisa, o CCLRC em Didcot (PENNOCK, 2006).

Em sua fase inicial, o DCC esteve mais voltado para um grupo-alvo definido, envolvendo as instituições de ensino superior e profissionais especialistas em dados, bem como bibliotecários, arquivistas, pesquisadores, entre outros. Com o tempo, DCC também incluiu os setores público e comercial em seu escopo, reconhecendo que o aprimoramento, tanto das ferramentas quanto dos processos para curadoria digital, dependiam de uma abordagem mais ampla, envolvendo também grupos de trabalho de organizações parceiras e padrões internacionais (DIGITAL CURATION CENTRE, 2004-2019a).

Assim, o desenvolvimento inicial da curadoria digital foi encabeçado pelo extenso trabalho do centro britânico, envolvendo uma comunidade internacional de pesquisadores e profissionais e a criação do evento internacional dedicado ao assunto, a *International Digital Curation Conference* (IDCC), bem como a criação do *International Journal of Digital Curation* (<http://www.ijdc.net>), periódico referência na temática.

Com o crescimento da importância dos estudos em curadoria digital, demais iniciativas institucionais foram estabelecendo-se pelo mundo com o apoio de centros de pesquisa e universidades. Destaca-se o trabalho interdisciplinar em curadoria digital da *Digital Curation Unit* (DCU), criada em 2007 como parte do *Athena Research Centre*, na

Grécia (DALLAS, 2007, CONSTANTOPOULOS; DALLAS, 2008, CONSTANTOPOULOS; DALLAS *et al.*, 2009) e o trabalho de educação profissional em curadoria digital encabeçado pelos Estados Unidos com a criação do projeto *International Digital Curation Curriculum* (DigCCurr) em 2006 (LEE; TIBBO; SCHAEFER, 2007, TIBBO; HANK, LEE, 2008).

Destarte, o desenvolvimento da curadoria digital se deu de modo diversificado, englobando diferentes perspectivas e objetivos, que vão desde interesses institucionais para a manutenção de recursos digitais até os estudos em educação e desenvolvimento acadêmico e de pesquisa. Percebe-se assim, que o campo tem estado em constante desenvolvimento desde sua proposta inicial, sendo também influenciado pelas mudanças tecnológicas que afetam todos os setores da atuação humana.

Harvey (2010, p.15, tradução nossa, grifo do autor) aponta para importantes fatores que caracterizam esse cenário de desenvolvimento:

- Imensas quantidades de informação em formato digital binário estão sendo geradas em todas as esferas da vida.
- As quantidades estão aumentando a uma taxa rápida.
- As comunidades científicas, acadêmicas e de pesquisa dependem cada vez mais da computação em rede, à medida que tendências como a mudança da ciência *in vitro* para *in silico* e o desenvolvimento de grandes bibliotecas digitais nas ciências humanas se tornam dominantes.
- A tecnologia de computadores (hardware, software e redes de comunicação) rapidamente se torna obsoleta.

Destaca-se que, estando orientada a esses fatores, a curadoria digital ainda encontra-se em construção e pode ser definida de diferentes formas dependendo da área de atuação ou de estudo. Sendo uma prática abrangente, ela é aplicável a uma grande variedade de situações e pode ser considerada como um termo guarda-chuva que está em desenvolvimento. Por isso, seu conceito também se encontra vinculado a diversas nomenclaturas e níveis de atuação, como ‘curadoria de informação’, ‘curadoria de conteúdo’, ‘curadoria de conhecimento’ e ‘curadoria de dados’ (JORENTE; SILVA; PIMENTA, 2015).

Nesse sentido, Constantopoulos *et al.* (2009, p.37, tradução nossa) destacam que

A prática da curadoria digital se estende a vários campos de atividade, abrangendo disciplinas de pesquisa das humanidades às ciências, bem como os produtos dessas disciplinas, quer sejam encontrados em repositórios de ciência eletrônica ou sob a custódia de gerentes de registros institucionais, ou em museus, bibliotecas e arquivos.

Em todos os casos, destaca-se que é a informação em meio digital, expressa pelos dados, que se torna preocupação central. Em outras palavras, é a capacidade de garantir que

os propósitos da informação em meio digital sejam atingidos e mantidos no futuro (CONSTANTOPOULOS; DALLAS, 2008), que coloca a curadoria digital como uma prática abrangente e discutida em vários âmbitos.

Tais propósitos estão relacionados com a capacidade dessas informações gerarem novos conhecimentos e servirem como base para novas pesquisas, o que vai ao encontro de Rusbridge *et al.* (2005) que consideram que o valor a longo prazo dos dados encontra-se no seu potencial como evidência e em suas possibilidades de reutilização, sendo a atividade de curadoria a chave para atingir esse potencial.

Por essa razão, Constantopoulos e Dallas (2008, p.1, tradução nossa) consideram que a

Curadoria digital é um novo campo interdisciplinar de investigação e comunidade de prática, que reúne as características disciplinares e os interesses de pesquisa de disciplinas como ciência de computação, arquivística, biblioteconomia e ciência da informação, disciplinas baseadas em coleções ou direcionadas a dados de pesquisa, como história da arte, arqueologia, biologia, ciências do espaço e da terra e áreas de aplicação, como repositórios de *e-science*, gerenciamento de registros organizacionais e, por último mas não menos importante, museus, bibliotecas e arquivos.

Ainda que se constitua como uma prática nova que só surgiu e ganhou destaque diante da abundância de informação em meio digital, a curadoria digital fundamenta-se na conceituação já dada a atividade de curadoria realizada em setores tradicionais. Esse conceito está relacionada com a etimologia da palavra, que tem sua origem no radical ‘*cur-*’ do termo latino ‘*cura(ae)*’. Segundo o Dicionário Houaiss (2009), tal termo remete à substantivos como cuidado e preocupação, direção e administração, curatela (jurídico), cuidado e tratamento (médico), guarda e vigia, objeto ou causa de cuidados ou preocupações e ainda, amor ou objeto amado. Também, do mesmo radical, origina-se o termo ‘*curator(óris)*’, que significa aquele que exerce a curadoria, o que está encarregado de algo, o curador.

De acordo com Beagrie (2006, p.4, tradução nossa), o termo ‘curadoria digital’, [...] beneficiou-se do uso já existente do termo “curadoria” usados pelos setores de bibliotecas e museus e pelas ciências biológicas.”, nos quais “[...] implica não apenas a preservação e manutenção de uma coleção ou banco de dados, mas algum grau de valor agregado e conhecimento.” Por isso, destaca-se que a curadoria digital construiu-se como um campo interdisciplinar que herda os aspectos tecnológicos, comunicacionais, gerenciais, cognitivos, de geração de conhecimento e informacionais, característicos da atividade de curadoria em ambientes tradicionais (SIEBRA; BORBA; MIRANDA, 2016).

Tal perspectiva vai ao encontro do pensamento de Glushko (2013), que entende que a atividade de curadoria ocorre em todos os sistemas de organização e pode ser realizada por

qualquer pessoa que tome decisões e emprega tecnologia para manter o conteúdo com qualidade e caráter ao longo do tempo. Por isso, exercer a curadoria, na visão de Glushko (2013, p. 194, tradução nossa), nada mais é do que definir “[...] políticas claras para coletar recursos e mantê-los ao longo do tempo, permitindo que as pessoas e os processos automatizados garantam que os dados dos recursos e suas representações sejam confiáveis, precisos, completos, consistentes e não redundantes.”

Diante disso, é importante verificar que o termo ‘curadoria’ associado ao termo ‘digital’ foi escolhido com a intenção de conceituar explicitamente as abordagens curatoriais direcionadas às coleções digitais, destacando a necessidade de mudanças em comparação com as abordagens empreendidas nos recursos analógicos (BEAGRIE, 2006).

A atividade de curadoria como é tradicionalmente entendida, denota um conjunto de atividades voltadas à seleção, cuidado e preservação de coleções de recursos informacionais, o que comumente “[...] se restringe contextos organizacionais relativamente limitados: bibliotecas, arquivos, museus, galerias de arte, herbários e instituições similares;” (NATIONAL RESEARCH COUNCIL, 2015, p.10, tradução nossa).

Em relação a curadoria digital, sua amplitude atinge uma grande diversidade de contextos organizacionais, não estando limitada a um único domínio específico, mas sim, estendendo-se a vários domínios nos quais os recursos digitais circulam e se relacionam.

Portanto, a curadoria digital apresenta certa continuidade com a atividade tradicional, contudo, leva em consideração principalmente os novos aspectos relacionados à informação, como a quantidade cada vez maior de recursos informacionais, a necessidade de gerenciamento ativo e contínuo diante da constante mudança de uso e tecnologias e ainda, a grande diversidade de contextos organizacionais (NATIONAL RESEARCH COUNCIL, 2015).

Percebe-se assim, que a curadoria tem suas atividades orientadas ao contexto que envolve a informação, incluindo toda a gama de questões que afetam desde sua criação e aquisição, processamento, manutenção, até o acesso e a disponibilização das informações. Para a curadoria digital, esse contexto se torna mais complexo, na medida em que os aspectos que envolvem a ciclo da informação também tomam proporções mais amplas.

Devido a essa complexidade, diversos autores tem tentado transmitir ao longo dos anos uma conceituação que defina os reais propósitos da curadoria digital. Essas definições são apresentadas no quadro 6.

Quadro 6 - Definições de curadoria digital

Autor	Definição
Beagrie (2004b, p. 7, tradução nossa)	O termo "curadoria digital" está sendo cada vez mais usado para as ações necessárias para manter dados de pesquisa digital e outros materiais digitais durante todo o seu ciclo de vida e ao longo do tempo para as gerações atuais e futuras de usuários.
Burnhill <i>apud</i> Giaretta (2004, não paginado, tradução nossa)	Curadoria digital: cuidar e, de alguma forma, "agregar valor" aos dados digitais. Isso provavelmente implica a criação de alguns dados novos a partir do existente, a fim de tornar o último mais útil e "adequado ao objetivo".
Rusbridge <i>et al.</i> (2005, p.32, tradução nossa)	[...] visão de curadoria abrange e vai além do aprimoramento da reutilização atual, e da responsabilidade de arquivamento, para abraçar a intensão que agrega valor através da provisão de contexto e ligação: colocando ênfase na publicação de dados de maneira a facilitar a reutilização e promover a responsabilização e integração.
Beagrie (2006, p.3, tradução nossa)	O termo curadoria digital está sendo cada vez mais usado para as ações necessárias para agregar valor e manter esses ativos digitais ao longo do tempo para as gerações atuais e futuras de usuários.
Pennock (2006, p.1, tradução nossa)	A curadoria digital, amplamente interpretada, é sobre a manutenção e agregação de valor a um corpo confiável de informações digitais para uso atual e futuro: em outras palavras, é o gerenciamento ativo e a avaliação da informação digital ao longo de todo o seu ciclo de vida.
Constantopoulos e Dallas (2008, p.3, tradução nossa)	De uma perspectiva do ciclo de vida da informação, a curadoria digital abrange uma série de processos voltados para alcançar (a) confiabilidade dos recursos digitais, (b) organização, arquivamento e preservação de longo prazo, e (c) serviços de valor agregado e novos usos para os recursos.
Abbott (2008, não paginado, tradução nossa)	Curadoria digital é a gestão e preservação de dados digitais a longo prazo. Todas as atividades envolvidas no gerenciamento de dados desde o planejamento de sua criação, melhores práticas na digitalização e documentação e a garantia de sua disponibilidade e adequação para descoberta e reutilização no futuro fazem parte da curadoria digital. A curadoria digital também pode incluir o gerenciamento de vastos conjuntos de dados para uso diário, por exemplo, garantindo que eles possam ser pesquisados e continuar legíveis.
Ball (2010, p.9, tradução nossa)	[...] a curadoria envolve gerenciar e promover o uso de dados desde seu ponto de criação. Ao planejar cuidadosamente a maneira como os dados são produzidos e processados, as tarefas posteriores de curadoria podem ser consideravelmente mais fáceis.
Yamaoka (2012, p. 5)	Um programa de curadoria digital tem como objetivo salvaguardar os objetos digitais, possibilitando o acesso e o reuso em todo seu ciclo de vida. O valor despendido para produzir tais objetos tem como retorno o compartilhamento dos dados, o que reduz a duplicação de esforços na criação destes e tornando-os disponíveis para extração de novos conhecimentos.
Sayão e Sales (2012, p.185)	A curadoria digital, em resumo, assegura a sustentabilidade dos dados para o futuro, não deixando, contudo, de conferir valor imediato a eles para os seus criadores e para os seus usuários. Os recursos estratégicos, metodológicos e as tecnologias envolvidas nas práticas da curadoria digital facilitam o acesso persistente a dados digitais confiáveis por meio da melhoria da qualidade desses dados, do seu contexto de pesquisa e da checagem de autenticidade.

Weidner e Alemneh (2013, não paginado, tradução nossa)	A curadoria digital é a atividade contínua de gerenciamento e aprimoramento do uso de recursos digitais ao longo de seu ciclo de vida e longo prazo.
Siebra <i>et al.</i> (2013, p. 9)	[...] envolve as atividades relacionadas à gestão dos dados, desde o planejamento da sua criação, passando pelas boas práticas na digitalização, na seleção dos formatos, na documentação e na garantia de estarem sempre disponíveis e adequados para serem descobertos e reusados agora e no futuro.
Santos (2014, p.137, grifo nosso)	A definição de curadoria digital como termo guarda-chuva para todas as ações voltadas à gestão da informação digital para preservação: inclui o planejamento de serviços e produtos, definição dos responsáveis pela informação nas instituições, levantamento das coleções e tipos de documentos a serem geridos, tratamento de documentos com <i>raw data</i> , como essa informação vai ser disponibilizada e acessada hoje e no futuro. A curadoria digital é um trabalho intervencionista, uma vez que se ocupa em agregar conteúdo à informação em meio digital afim que de enriquecê-lo para preservá-lo.
Clatin <i>et al.</i> (2014, p.1, tradução nossa)	A curadoria digital é o conceito global que inclui todos os aspectos do trabalho sobre os bens culturais digitais, digitalizados e os nascidos digitais: desde a entrada do documento, à descrição dos dados, armazenamento, disseminação e preservação a longo prazo.
National Research Council (2015, p.1, tradução nossa)	[...] a curadoria digital envolve mais do que o armazenamento seguro e a preservação da informação digital, porque a curadoria pode agregar valor à informação digital e aumentar sua utilidade. [...] O que distingue a curadoria desses outros campos é sua ênfase em aumentar o valor dos ativos de informação para uso atual e futuro e sua atenção para o reaproveitamento e a reutilização de informações, tanto dentro quanto fora do contexto em que foi criada ou coletada.
Oliver e Harvey (2016, p.8, tradução nossa)	A curadoria digital é um conceito mais inclusivo do que o arquivamento digital ou a preservação digital. Aborda toda a gama de processos aplicados a objetos digitais ao longo de seu ciclo de vida. A curadoria digital começa antes que os objetos digitais sejam criados, definindo padrões para o planejamento da coleta de dados que resulta em objetos digitais "preparados para curadoria" que estão na melhor condição possível para garantir que possam ser mantidos e usados no futuro. A curadoria digital enfatiza a adição de valor a conjuntos de dados e objetos digitais, por meio de itens como metadados ou anotações adicionais, para que possam ser reutilizados.
Poole (2016, p. 961, tradução nossa)	[...] a curadoria digital também agrega valor aos ativos de dados confiáveis para uso atual e futuro.
Weber (2017, p.144)	[...] o estado atual do conceito de curadoria digital, no periódico <i>International Journal of Digital Curation</i> , está associado a outros termos como: curadoria de dados, curadoria de dados digitais, e acima de tudo está atrelado ao ciclo da preservação digital, ao reuso de dados científicos, ao fornecimento de metadados de alta qualidade, ao compartilhamento de dados e a melhoria da ciência.
Higgins (2018, p. 13, tradução nossa)	A curadoria digital é uma disciplina que se preocupa com um objeto de pesquisa principal - dados (aqui definidos como informações que se manifestam digitalmente) e um único paradigma - garantir o acesso contínuo e o uso de dados a longo prazo.
Digital Curation Centre (2004-2019d, não paginado, tradução nossa)	Curadoria digital é manter e agregar valor a um corpo confiável de dados de pesquisa digital para uso atual e futuro; engloba o gerenciamento ativo de dados durante todo o ciclo de vida da pesquisa.

Fonte: Elaborado pela autora.

A partir da análise das definições sobre curadoria digital apresentadas na literatura é possível destacar algumas interpretações.

Em primeiro lugar, que o objeto central da curadoria digital são os dados. Sendo que esses dados ou conjunto de dados podem ser considerados como objetos digitais, englobando tudo aquilo é armazenado como fluxos de *bits*, ou como informação digital que se refere às informações que se manifestam digitalmente. Também nota-se o uso da palavra ativos digitais ou ativos de informação para se referir ao objeto da curadoria digital, indicando um conteúdo em formato binário que inclua os direitos de uso.

Assim, compreende-se que o objeto da curadoria digital são os recursos informacionais expressos na forma digital, ou seja, se referindo a itens que sejam dotados de contexto e relações de significado e passíveis de interpretação e descrição, podendo ser mais simples, como a expressão mínima de um conteúdo (dados), ou mais complexos, envolvendo grandes conjuntos de dados.

De acordo com Santos, Simionato e Arakaki (2014),

O termo recurso informacional refere-se a informação objetivada no contexto de um campo do conhecimento podendo ser apresentado em uma estrutura analógica e/ou digital, com valor informacional que caracteriza a sua concepção intelectual expressa na corporificação de manifestações estruturadas na forma de itens.

Sendo assim, a segunda interpretação refere-se ao fato de que a curadoria digital é uma atividade que é direcionada a garantir o valor informacional dos recursos, ou seja, que tem direta relação com a sua concepção intelectual. Isso implica não só nas ações de arquivamento e preservação digital, mas principalmente em manter e agregar valor por meio da representação da informação para que esses recursos possam atender aos propósitos informacionais e fomentar novos conhecimentos.

Portanto, destaca-se que o objetivo da curadoria digital é garantir o valor dos recursos informacionais para uso atual e futuro, para que eles possam não só se manter acessíveis, mas com qualidade e confiabilidade para serem reaproveitados e reutilizados em diversos contextos.

Nessa perspectiva, percebe-se que o objetivo da curadoria digital não é algo simples, mas sim que se ramifica em diversos objetivos para atingir um objetivo total. Por isso sua abordagem não é linear, e sim, cíclica, levando em consideração todos os estágios de vida e as relações de um recurso informacional. Para isso a curadoria digital se vale do emprego de estratégias, de métodos e de tecnologias para proporcionar esses objetivos, o que exige a habilidade de diversos profissionais qualificados.

O emprego das práticas de curadoria digital em diferentes estágios de manejo dos recursos digitais podem influenciar a capacidade de usá-lo e por isso “[...] uma abordagem de ciclo de vida garante que todas as etapas necessárias sejam identificadas e planejadas e que as ações necessárias sejam implementadas na sequência correta.” (HIGGINS, 2008, p. 135, tradução nossa). De acordo com Pennock (2006, p. 3, tradução nossa), uma abordagem do ciclo de vida é necessária porque:

- Os materiais digitais são frágeis e suscetíveis de mudar de avanços tecnológicos ao longo de seu ciclo de vida, ou seja, da criação em diante;
- Atividades (ou falta de) em cada estágio do ciclo de vida influenciam diretamente nossa capacidade de gerenciar e preservar materiais digitais em etapas subsequentes;
- A reutilização confiável de materiais digitais só é possível se os materiais forem curados de forma que sua autenticidade e integridade sejam mantidas

Como entende Santos (2014), a curadoria digital é um empreendimento que exige coordenação de recursos humanos e operacionais, que por sua vez apontam para a execução de uma sequência de procedimentos que devem ser aplicados aos recursos digitais, de acordo com algum modelo.

Ao longo dos anos, diversas abordagens em forma de modelos gráficos foram desenvolvidas para o ciclo de vida de curadoria de recursos digitais. Algumas dessas deram origem a modelos de ciclo de vida com estruturas específicas, outras, à práticas voltadas para vários domínio de atuação, como será visto a seguir.

3.1 Os modelos de curadoria digital

Os modelos de ciclo de vida desenvolvidos para orientar as ações de curadoria digital ganham destaque diante da grande quantidade de dados que podem ser acessados por meio de ambientes digitais, tais como repositórios, bibliotecas digitais, museus virtuais, plataformas digitais, entre outros.

Desde o princípio, as diferentes partes interessadas na gestão de dados viram a importância de uma abordagem em ciclo de vida para a manutenção digital (BEAGRIE, 2006).

Um grande passo nesse sentido foi o entendimento de que a noção de preservação digital deveria acompanhar “[...] a percepção de que a perpetuação de recursos digitais a longo prazo envolve a observância de práticas cuidadosas de gerenciamento de ativos digitais difundidas ao longo do tempo.” (LAVOIE; DEMPSEY, 2004, não paginado, tradução nossa). Assim, o foco da preservação digital deixou de ser direcionado para as medidas imediatas e

pontuais e voltou-se para os processos distribuídos ao longo do ciclo de vida da informação digital, operando em conjunto com todos os serviços que suportam os ambientes digitais (LAVOIE; DEMPSEY, 2004).

Segundo Santos (2014, p. 59), os modelos de ciclo de vida da informação “[...] podem ser definidos como o processo como a informação em meio digital se move através de estágios, desde a sua criação até a preservação contínua, gestão e acesso ao longo do tempo”. Baseados nesses estágios pelos quais a informação digital se move, uma variedade de propostas de ciclo de vida de curadoria foram desenvolvidos, resultando em uma variedade de modelos que delineiam a estrutura de operações a serem executadas para recursos informacionais durante toda a sua vida (BALL, 2012).

Grande parte dos modelos criados se destina a atender às necessidades de comunidades específicas, em especial, os modelos direcionados à curadoria de dados científicos e de pesquisas em *e-Science*, que são representativos nos estudos da área, por exemplo, o modelo do *DataONE*, que estabelece recomendações sobre como trabalhar efetivamente com os dados em todos os estágios do ciclo de vida (DATAONE, 2019) e o modelo do *UK Data Archive Data Lifecycle*, que aborda o gerenciamento de dados e metadados para que permaneçam disponíveis e compreensíveis a longo prazo tendo em vista o avanço da pesquisa científica (UK DATA ARCHIVE, 2002-2019). Porém existem aqueles que se configuram como padrões e especificações próprios para a prática de curadoria digital.

Nessa perspectiva, pode-se considerar que a curadoria digital é um conceito amplo com uma abordagem que reúne toda a gama de processos aplicados aos recursos digitais ao longo de seu ciclo de vida levando em consideração as necessidades dos usuários e dos sistemas.

A seguir, são apresentados quatro modelos de ciclo de vida de curadoria digital que possuem relevância para o propósito de delinear a prática de curadoria digital. Destaca-se o modelo de referência OAIS, o modelo do JISC, o modelo do *Digital Curation Centre* (DCC) e o modelo o *Extended Digital Curation Lifecycle Model* do DCU (DCC&U).

3.1.1 Modelo de referência OAIS

Um importante modelo, denominado *Open Archival Information Systems* (OAIS) *Reference Model*, foi idealizado na década de 90, tendo em vista a nova abordagem para a preservação digital. Segundo Ferreira (2006), o modelo OAIS tem origem no esforço conjunto entre o *Consultative Committee for Space Data Systems* (CCSDS) e o *International*

Organization for Standardization (ISO) para desenvolver normas capazes de regular o armazenamento a longo-prazo de informação digital de dados espaciais.

Posteriormente, o modelo foi amplamente recomendado para gerenciar recursos em sistemas de arquivamento digital, fornecendo uma estrutura genérica para a orientação e implementação de um sistema de arquivamento digital. Em 2003, o modelo foi aprovado como uma norma internacional, sob a ISO 14721:2003, ganhando ampla aceitação. (HIGGINS; SEMPLE, 2006; HARVEY, 2010).

Conforme Thomaz e Soares (2004, p.8), o modelo “[...] é um esquema conceitual que disciplina e orienta um sistema para a preservação e manutenção do acesso à informação digital por longo prazo” Nessa perspectiva, Harvey (2010) aponta para as três principais características do modelo OAIS:

Em primeiro lugar, ele fornece um vocabulário de conceitos relacionados à preservação que é compreendido e adotado por pessoas de diversos domínios (bibliotecários, arquivistas, profissionais de patrimônio cultural, pessoal de tecnologia da informação [TI], cientistas, acadêmicos). Em seguida, define um modelo de informação. Em terceiro lugar, define um modelo funcional: isto é, descreve as principais funções necessárias em um arquivo digital e fornece informações sobre os tipos de atividades realizadas por cada função.

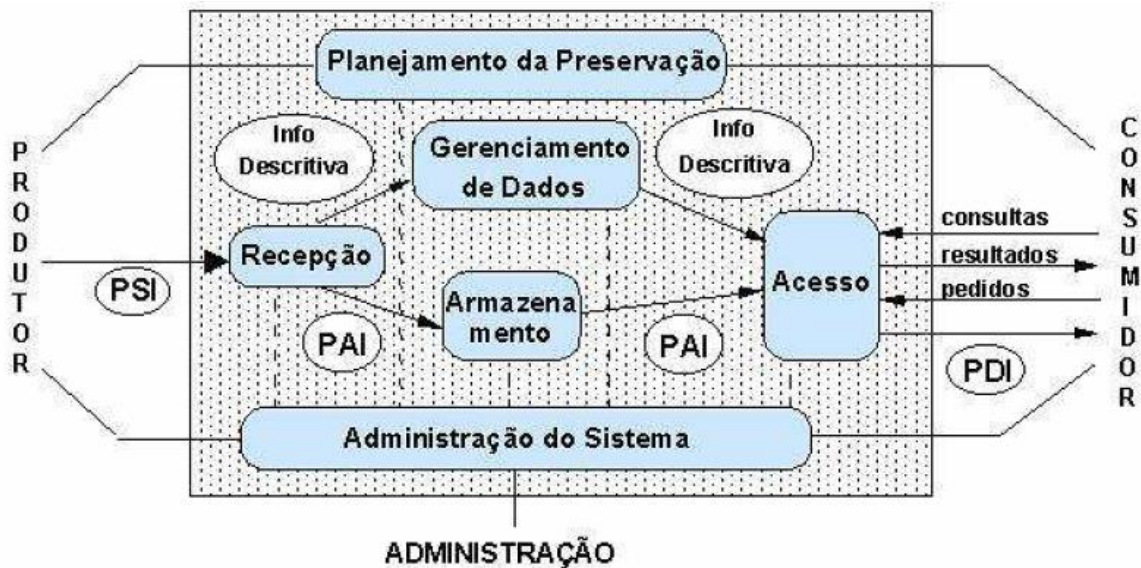
A primeira característica se refere ao ambiente OAIS, que é o contexto em que se insere um sistema de arquivamento, bem como as entidades que participam do ambiente e suas interações. A segunda característica refere-se ao modelo de informação OAIS que consiste na atribuição de metadados responsáveis por registrar os processos associados à preservação, estabelecer a autenticidade dos recursos digitais, bem como especificar direitos de acesso e identificar de forma única. Nessa parte são identificados quatro tipos de informações, que são chamadas de objetos de informação. São elas, a Informação de Conteúdo, Informação de Descrição de Preservação, Informação de Pacote e Informação Descritiva. Por fim, a terceira característica refere-se à explicitação funcional do modelo, que descreve as principais funções a serem realizadas em um sistema de arquivamento, reunido as características anteriores em um modelo funcional (THOMAZ; SOARES, 2004).

Desse modo, ressalta-se que a função do modelo não é especificar uma dada implementação, mas sim fornecer uma estrutura conceitual para planejar uma implementação efetiva, descrevendo as funcionalidades básicas necessárias para um sistema de arquivamento e definindo as responsabilidades e as interações entre os agentes Produtores, Consumidores e Administração do Sistema. Por essa razão, ele se configura como um padrão que define os processos necessários para a preservação efetiva e o acesso a longo prazo, ao mesmo tempo

em que estabelece uma linguagem comum e um método padronizado para descrever a funcionalidade do sistema de arquivamento (HIGGINS; SEMPLE, 2006).

O modelo de referência OAIS estabelece um modelo funcional detalhado com informações e funções de arquivamento, como possível visualizar na figura 3.

Figura 3 - Modelo funcional OAIS



Fonte: Thomaz e Soares (2004).

Uma importante característica do modelo OAIS é o fato de ele ser baseado no conceito de Informação de Pacote. Esses pacotes carregam as informações associadas aos recursos digitais. Segundo Thomaz e Soares (2004, p.11), uma Informação de Pacote “[...] reúne a Informação de Conteúdo e a Informação de Descrição de Preservação em um pacote identificável, enquanto a Informação Descritiva facilita o acesso à Informação de Pacote através de ferramentas de pesquisa e recuperação.” O modelo funcional OAIS possui seis entidades funcionais. São elas: *Recepção*, *Armazenamento*, *Gerenciamento de Dados*, *Administração do Sistema*, *Planejamento de Preservação* e *Acesso* (THOMAZ; SOARES, 2004). De acordo com Thomaz e Soares (2004) as entidades são definidas como:

Recepção: aceite e preparação dos *Pacotes de Submissão de Informação* (PSI) para armazenamento e gerenciamento dentro do arquivo, gerando a *Informação Descritiva* e os *Pacotes de Informação de Arquivamento* (PAI) e transferindo-os para a entidade *Armazenamento* e para a entidade *Gerenciamento de Dados*, respectivamente.

Armazenamento: armazenamento, manutenção e recuperação de PAI, envolvendo receber novos PAI da entidade *Recepção* e posicioná-los na área de armazenamento permanente de acordo com diversos critérios (requisitos de suporte, taxas de utilização esperada, etc.), fornecendo cópias de PAI solicitados à entidade *Acesso*.

Gerenciamento de Dados: mantém e acessa tanto a *Informação Descritiva*, que identifica e documenta os recursos, quanto os dados administrativos usados para gerenciá-lo, administrando a base de dados do sistema, promovendo as atualizações e consultando os dados da entidade para gerar relatórios.

Administração do Sistema: gerencia a rotina operacional de todo o sistema, analisando o desempenho geral das operações do arquivo e as operações de migração e atualização de seus conteúdos, sendo responsável por cumprir e manter os padrões e políticas e fornece suporte ao cliente.

Planejamento de Preservação: monitora o sistema e fornece recomendações para garantir que a informação armazenada permaneça acessível por longo prazo, avaliando o conteúdo, recomendando padrões e políticas do arquivo e monitorando as mudanças no ambiente tecnológico e nas demandas de serviço. Ocorre o desenvolvimento de pacotes de informação (PAI e PSI) para submissões específicas.

Acesso: viabiliza a comunicação com os *Consumidores* no momento da determinação da existência, descrição, localização e disponibilidade da informação armazenada no sistema, permitindo que os *Consumidores* solicitem e recebam produtos de informação. Nessa etapa ocorre o desenvolvimento de Pacotes de Disseminação de Informação (PDI), recebendo pedidos, entregando resultados e gerando relatórios.

Desse modo, o modelo de referência OAIS permite gerenciar o fluxo de informações internas que compõe um sistema de arquivamento, viabilizando a preservação e o acesso a longo prazo a partir das informação atribuídas. Devido sua aplicabilidade para o arquivamento de diversos tipos de recursos, o modelo OAIS tornou-se amplamente adotado pelas comunidades digitais (BALL, 2010), sendo a base para vários outros modelos, apoiados por iniciativas nacionais e internacionais e voltados para diferentes domínios.

3.1.2 Modelo do JISC

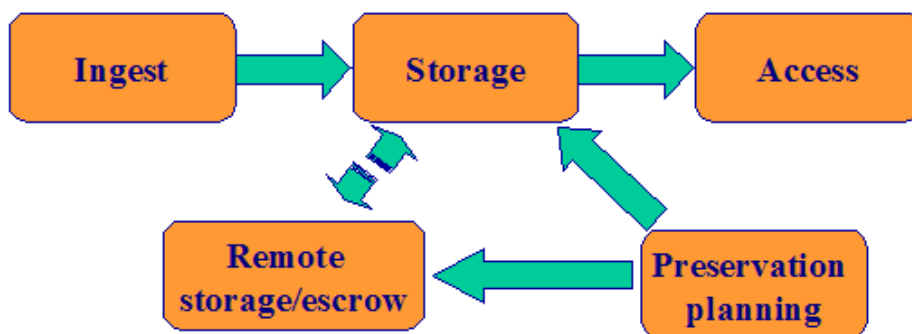
No setor de educação e pesquisa do Reino Unido, a organização britânica sem fins lucrativos, *Joint Information Systems Committee* (JISC) teve grande engajamento nesse sentido, publicando em 2002 o *Continuing Access and Digital Preservation Strategy for the*

Joint Information Systems Committee (JISC) 2002-5, um plano estratégico que propunha um maior investimento no setor de gestão e preservação digital.

Diante dos objetivos do plano estratégico, o comitê JISC investiu fortemente no uso e desenvolvimento de arquiteturas, padrões e práticas adotadas para criação e gerenciamento de recursos digitais. Com resultado, a organização envolveu-se no desenvolvimento de estruturas para o gerenciamento do ciclo de vida dos dados, entre elas o modelo OAIS. No entanto, como explica Beagrie (2004a), o JICS tinha uma preocupação mais ampla acerca do processo de preservação, envolvendo além do sistema interno de arquivamento digital, os serviços de hospedagem e de acesso, bem como as coleções digitais e os serviços nacionais distribuídos. Por essa razão, algumas adaptações foram necessárias em relação ao modelo OAIS, devido ao fato dele levar em conta apenas as atividades dentro do sistema de arquivamento (HARVEY, 2010).

Assim, um outro modelo foi desenvolvido com o foco no planejamento do armazenamento e da preservação em um ambiente mais complexo, com retenção distribuída e preservação ao longo de períodos variáveis, dando origem ao *Joint Information Systems Committee Model*, um modelo de ciclo de vida para sistemas de informação JISC (BEAGRIE, 2004). Na figura 4, é possível visualizar uma versão adaptada e simplificada do modelo que é apresentada no plano estratégico do JISC.

Figura 4 - Modelo simplificado para o ambiente informacional JISC



Fonte: Beagrie (2004a).

Tal como é apresentado na figura 4, o modelo do JISC leva em consideração aspectos sequenciais e interdependentes, desde a ingestão de dados, armazenamento seguro e externo, planejamento da preservação, até a garantia de acesso aos dados. Como destaca Silva (2017, p. 40) “O modelo retrata a inquietação do JISC com os objetos digitais que serão inseridos, arquivados, salvaguardados e disponibilizados em um sistema.”

Tais inquietações levam o JISC a defender em seu plano estratégico, a criação de um centro de curadoria digital nacional, o *Digital Curation Centre* – DCC, como o objetivo de fornecer serviços de curadoria e preservação e promover o desenvolvimento de novas pesquisas no setor. (BEAGRIE, 2004).

É nesse cenário que a preservação digital passa por uma mudança de paradigma, deixando de ser vista como algo isolado e tornando-se parte dos processos da curadoria digital. Desse modo, como apontam Siebra *et al.* (2013), a curadoria digital se torna mais ampla que a preservação digital, envolvendo uma série de procedimentos que vão do planejamento da criação dos dados até a garantia de seu acesso e uso.

Para Conway (2001), os conceitos atribuídos à preservação tradicional são transformados pelo universo digital, retirando o foco da integridade física do objeto e passando a destacar a integridade intelectual como característica principal. Nessa mesma perspectiva, Sayão e Sales (2012, p. 180), entendem que o principal desafio desse cenário em que surge a curadoria digital é a “[...] necessidade de se preservar não somente o conjunto de dados, mas de preservar, sobretudo, a capacidade que ele possui de transmitir conhecimento para uso futuro das comunidades interessadas.”

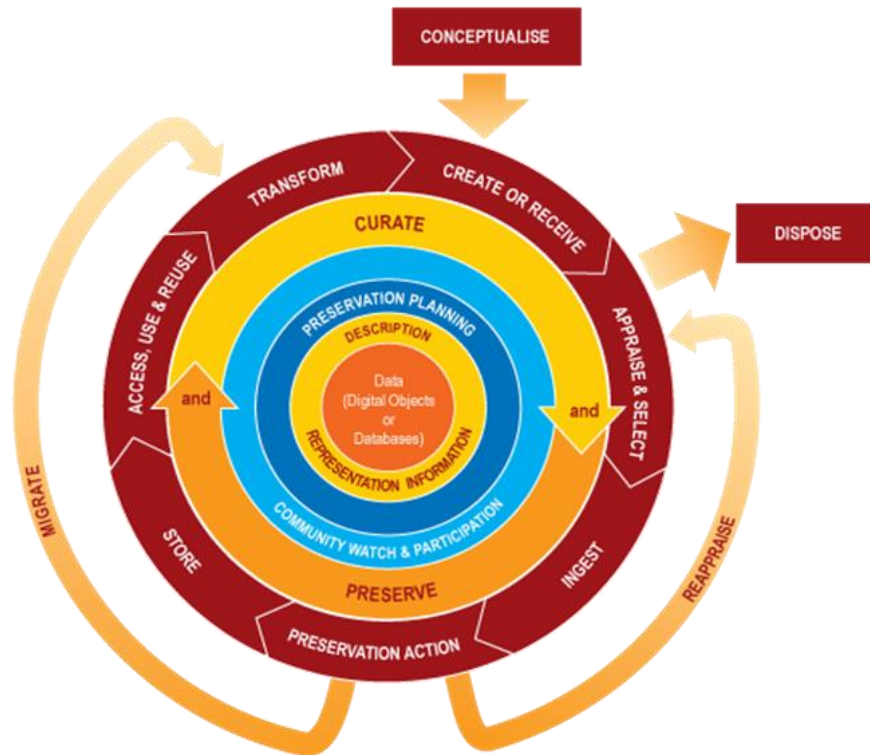
Nesse sentido, destaca-se que a curadoria digital está preocupada com todas as instâncias que se relacionam a manutenção do valor informacional de um recurso digital. Com esse foco, *Digital Curation Centre* (DCC) desenvolveu um modelo próprio que propõe que as todas as ações da curadoria digital giram em torno dos recursos digitais em formato de ciclo de vida.

3.1.3 Modelo de ciclo de vida de curadoria do DCC

O *Digital Curation Centre* (DCC) desenvolveu um modelo com a intenção de contemplar todos estágios necessários para uma curadoria e preservação bem-sucedidas, envolvendo desde a conceituação inicial ou recebimento dos dados até um ciclo de curadoria iterativo (DIGITAL CURATION CENTRE, 2004-2019b). Denominado *DCC Curation Lifecycle Model*, o modelo foi projetado para abordar especificamente as necessidades da curadoria digital (HIGGINS, 2008) sendo sua primeira versão publicada em 2007, e após um período de consulta pública, finalizado em 2008. (HARVEY, 2010).

Como apresentado na figura 5, o modelo fornece uma visão geral gráfica de alto nível dos estágios necessários para empreender a curadoria digital (HIGGINS, 2008).

Figura 5 - DCC *Digital Curation Lifecycle Model*



Fonte: Higgins (2008).

O modelo é uma abstração que permite planejar as atividades de curadoria digital em diferentes ambientes digitais e por isso é aplicável a uma grande variedade de situações e implica em diversos procedimentos interdependentes e relacionados para atingir sua finalidade.

Segundo Higgins (2007a, p.83, tradução nossa):

O modelo pode ser usado para planejar atividades de curadoria e preservação dentro de uma organização ou consórcio para garantir que todas as etapas necessárias sejam realizadas, cada uma na sequência correta. O modelo permite que a funcionalidade granular seja mapeada para definir funções e responsabilidades e construir uma estrutura de padrões e tecnologias para implementar.

Nesse sentido, no centro do ciclo encontram-se os dados, também chamados de objetos digitais, são qualquer elemento em formato binário processável por máquina. Isso inclui: os objetos digitais simples, que correspondem a arquivos de som, textuais ou de imagens, juntamente com seus identificadores e metadados relacionados; os objetos digitais complexos, que consistem na combinação de vários outros objetos digitais, como *websites*. E

por fim, as bases de dados, que são coleções estruturadas de registros ou dados armazenados em um sistema (HIGGINS, 2008).

As ações que compõem o modelo de ciclo de vida de Curadoria Digital do DCC são divididas em **Ações para Todo o Ciclo de Vida**; **Ações Sequenciais**; e **Ações Ocasionais**, para abranger as necessidades de gestão dos recursos digitais. Com base no estudo de Higgins (2008) apresentam-se as ações e a descrição dos seus respectivos processos:

a) Ações para Todo o Ciclo de Vida - Ações que permeiam todo o ciclo de vida da curadoria digital envolvendo os objetos digitais que estão no centro do modelo.

- *Descrição e representação da informação*: atribuição de metadados (administrativos, descritivos, técnicos, estruturais e de preservação), utilizando os padrões de metadados adequados para assegurar uma descrição e controle a longo prazo. Coletar e atribuir informações de representação necessárias para entender e processar o material digital e os metadados associados;

- *Planejamento da preservação*: plano de preservação ao longo do ciclo de vida do material digital. Isso incluiria planos para gerenciamento e administração de todas as ações do ciclo de vida de Curadoria;

- *Acompanhamento e participação da comunidade*: manter um plano de controle sobre atividades comunitárias apropriadas e participar do desenvolvimento de padrões compartilhados, ferramentas e *softwares* adequados;

- *Curadoria e preservação*: ações administrativas e de gestão planejadas para promover a conservação e a conservação ao longo do ciclo de vida da cura;

b) Ações Sequenciais – Ações realizadas repetidamente para assegurar que objetos digitais permaneçam em contínuo processo de manutenção.

- *Conceptualização*: planejar a criação de dados, incluindo o método de captura e opções de armazenamento.

- *Criação e/ou Recepção*: criação de dados incluindo metadados administrativos, descritivo, técnico, estrutural e de preservação; ou Recepção de dados (dados já criados por outros arquivos, repositórios ou *data centers*) de acordo com políticas de coleta de dados, atribuindo metadados mais apropriados caso necessário;

- *Avaliação e seleção*: avaliação e seleção dos dados para curadoria e preservação de longo prazo. Ser aderente às orientações documentadas, políticas e exigências legais;

- *Admissão*: transferir dados de um arquivo, repositório, *data center* ou outra entidade custodiadora, seguindo as políticas e os requisitos legais;
- *Ação de preservação*: ações para a preservação que devem garantir que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo sua integridade. Tais ações incluem limpeza de dados, validação, atribuição de metadados de preservação e representação e garantia de estruturas ou formatos de arquivos dados adequados.
- *Armazenamento*: armazenamento dos dados de forma segura, aderindo aos padrões relevantes;
- *Acesso, uso e reuso*: viabilizar acesso aos dados para uso e reuso pelo usuário, tanto pela publicação de forma aberta, como com controle de acesso e procedimentos de autenticação;
- *Transformação*: criação de novos dados a partir do original (migração para formatos diferentes ou criação de subconjuntos, por seleção ou consulta, para criar resultados derivados);

c) Ações Ocasionais – Ações eventuais realizadas sobre as ações sequenciais em decorrência de uma decisão ou mudança no curso do processo.

- *Descarte*: eliminar dados que não foram selecionados para a curadoria e preservação, de acordo com as políticas documentadas, orientação ou requisitos legais estabelecidos. Esses dados podem ser transferidos para outro local de guarda ou destruídos, atendendo aos requisitos legais para uma destruição segura, mediante a natureza dos dados;
- *Reavaliação*: Reavaliação e nova seleção dos dados que retornam por falhas nos procedimentos de validação;
- *Migração*: migração de dados para outro formato visando garantir a imunidade dos dados à obsolescência de *hardware* ou *software* ou para adequação com o ambiente de armazenamento.

Desse modo, cada um dos processos é detalhadamente descrito para orientar a tomada de decisão em relação ao fluxo de dados e em relação as próprias ações. A partir do modelo, é possível visualizar as ações que são, ou que deixam de ser, aplicadas em cada estágio do ciclo de vida dos dados, influenciando diretamente na eficácia com que as informações podem ser gerenciadas e preservadas nos estágios seguintes do ciclo de vida (HARVEY, 2010).

Diante disso, destaca-se que essa interdependência entre ações é a grande chave para empreender a curadoria digital. Como explica Harvey (2010), as Ações para Todo o Ciclo de Vida se aplicam à maioria das Ações Sequenciais: as atividades de Planejamento de preservação são atividades contínuas que devem ser levadas em conta à medida que os dados são movidos pelas Ações Sequenciais e são especialmente relevantes para ‘Conceitualização’, ‘Ação de preservação’ e ‘Acesso, uso e reutilização’.

Do mesmo modo, é possível destacar o encadeamento das funções das Ações Sequenciais com as Ações Ocasiais. Um objeto digital, após ser, criado ou recebido (Criação e Recepção), avaliado e selecionado (Avaliação e Seleção) e admitido (Admissão), passa para as ações de preservação, caso ele atenda aos requisitos, ele passara para a próxima ação sequencial, que é o ‘Armazenamento’, e seguirá pelo processo. No entanto, caso não esteja em conformidade com os padrões de preservação, ele poderá passar por dois outros estágios que são ocasionais, decorrentes de alguma modificação no fluxo normal do processo.

Na primeira possibilidade, o objeto digital passará para a ação de ‘Reavaliação’ (Ações Ocasiais), e retornará para a ação de ‘Avaliação e Seleção’ (Ações Sequenciais). Nesse caso, se o objeto digital for avaliado e selecionado de modo satisfatório, ele continuará no processo sequencial. Em caso não satisfatório, ele será enviado para o ‘Descarte’ e será eliminado atendendo às regras.

A outra possibilidade é a ação de ‘Migração’. Nesse caso, se a ação de preservação exigir alguma modificação no objeto digital para atender aos padrões de preservação, então ocorrerá a migração para formatos diferentes, incidindo na ação de ‘Transformação’ e resultando em um novo objeto digital.

Diante disso, evidencia-se que o modelo do DCC foi desenvolvido para alinhar as práticas de curadoria digital com os estágios do ciclo de vida dos dados (BALL, 2010). Por isso, uma efetiva curadoria digital depende muito das informações atribuídas ou associadas aos dados. A atribuição ou associação de informações ocorrem em todos os estágios do ciclo de vida dos dados, pois não é possível garantir sua preservação e que eles sejam utilizáveis continuamente sem essas informações (HARVEY, 2010).

O modelo de ciclo de vida do DCC refere-se a essas informações atribuídas ou associadas como informações de descrição e de representação. As principais atividades associadas às ações de descrição e de representação no Ciclo de vida de Curadoria do DCC são:

- Valorizar a necessidade de informações de descrição e representação;
- Estar ciente de onde as informações de descrição e de representação são necessárias;
- Compreender os principais padrões que existem para a descrição e a representação de informações; e
- Desenvolver políticas para aplicação de informações de descrição e de representação. (HARVEY, 2010, p. 66, tradução nossa).

Nesse contexto, destaca-se a importância função da ação de *Descrição e de Representação da Informação* que compõe as *Ações para Todo o Ciclo de Vida*, como responsável pela definição de metadados em todos os estágios do ciclo de curadoria digital.

Tais metadados devem possibilitar:

- Identificar de forma persistente os objetos digitais;
- Manter *links* confiáveis para os objetos digitais;
- Descrever claramente o que são os objetos digitais;
- Identificar claramente as características técnicas dos dados que compõem os objetos digitais;
- Identificar quem é responsável pela gestão e preservação dos objetos digitais;
- Descrever o que pode ser feito para os objetos digitais;
- Descrever o que é necessário para representar os objetos digitais no padrão exigido pelos usuários;
- Registrar a história dos objetos digitais; e
- Documentar a autenticidade dos objetos digitais. (HARVEY, 2010, p.65, tradução nossa).

O modelo do DCC faz distinção entre as informações chamadas de ‘informações de representação’ e as chamadas ‘informações de descrição’. Porém, ambas são consideradas informações que permitem a representação dos dados.

As informações de descrição tem a finalidade de garantir a descrição e o controle adequados a longo prazo, envolvendo os diferentes tipos de metadados (HARVEY, 2010). Harvey (2010, p.65, tradução nossa) pontua quais são as informações de descrição para fins da atividade de curadoria:

- Metadados que descrevem objetos digitais e sua localização (os metadados descritivos e os metadados estruturais observados no Modelo de Curadoria de Ciclo de Vida)
- Metadados que fornecem as informações técnicas necessárias para usar objetos digitais (os metadados técnicos observados no Modelo)
- Metadados que descrevem o que aconteceu com os objetos digitais à medida que eles se movem no ciclo de vida da curadoria (os metadados administrativos e os metadados de preservação no Modelo).

Já as informações de representação são atribuídas para entender e processar os dados. Esse tipo de informações, que também são entendidas como metadados, devem explicitar a sintaxe e a semântica dos dados para que uma comunidade específica possa acessar as

informações armazenadas em um objeto digital. São exemplos de informações de representação os vocabulários de dados, as ontologias e os tesouros, que fornecem informações para definir os relacionamentos entre os objetos e suas partes. (HARVEY, 2010).

Segundo a definição do *Digital Curation Centre* (2004-2019c, não paginado, tradução nossa):

Informações de representação são todas as informações necessárias para entender e processar o material digital e os metadados associados. Objetos digitais são armazenados como fluxos de bits, que não são compreensíveis para um ser humano sem dados adicionais para interpretá-los. A informação de representação é a informação extra estrutural ou semântica, que converte dados brutos em algo mais significativo.

Desse modo, percebe-se que o modelo reflete as principais preocupações do DCC em relação aos objetos digitais, envolvendo “Questões de autenticidade e integridade, estratégias para fornecer representação e acesso adequados ao conhecimento, suporte para um ciclo de vida previsível de preservação de ativos digitais, bem como atenção aos interesses de comunidades particulares [...]” (CONSTANTOPOULOS; DALLAS *et al.*, 2009, p.35, tradução nossa).

Tal como um modelo genérico, o modelo de DCC prevê uma visão geral dos estágios necessários para empreender a curadoria digital, o que explica sua ampla aplicabilidade, atravessando diversos contextos e disciplinas. No entanto, de modo paralelo a ideia de curadoria digital como um campo universal, desenvolveu-se um conjunto de estudos voltados ao domínio dos patrimônios culturais que defendia a necessidade de aprofundamento das práticas de curadoria digital direcionadas a disciplinas e contextos específicos (CONSTANTOPOULOS; DALLAS *et al.*, 2009).

Um conjunto desses estudos resultou na formulação da proposta de Constantopoulos e Dallas (2008) para um modelo de ciclo de vida estendido, o *Extended Digital Curation Lifecycle Model* (DCC&U), que tem por base as diretrizes estabelecidas pelo DCC, mas com o acréscimo da noção de informação contextual e aprimoramento do conhecimento para os recursos digitais.

3.1.4 Modelo de ciclo de vida de curadoria estendido - DCC&U

O *Extended Digital Curation Lifecycle Model* (DCC&U) foi desenvolvido como resultado da percepção de que a curadoria digital precisava abranger não só os recursos digitais em si, mas também a capacidade de adequação epistêmica da informação em diversos

contextos de uso, de modo a assegurar a validade e a utilidade dos recursos digitais para a adequação ao seu propósito informacional. Tal abordagem teve como base o extenso trabalho em informática voltada para patrimônios culturais realizado pelo *Digital Curation Unit* (DCU), no *Athena Research Centre*, em Atenas, na Grécia (DALLAS, 2007, CONSTANTOPOULOS; DALLAS, 2008, CONSTANTOPOULOS; DALLAS *et al.*, 2009).

Tendo como base o modelo do *Digital Curation Centre* (DCC), essa abordagem alternativa proposta pelos pesquisadores da DCU leva em consideração novas estratégias importantes. A primeira estratégia se refere a uma abordagem de ciclo de vida que inclua no processo de curadoria digital, as representações dos recursos digitais, permitindo sua contínua atualização e enriquecimento em conformidade com as comunidades designadas. A segunda estratégia sugere o uso de metodologias centradas em eventos para abordar os estágios do recurso digital no seu ciclo de vida. A próxima estratégia diz respeito a ampliação do atores envolvidos, reconhecendo que estes incluem não apenas os guardiões de recursos, como bibliotecários e gestores de dados, mas também aqueles preocupados com a produção de conhecimento gerados por esses recursos e as comunidades de usuários. A última estratégia reconhece a necessidade de levar em consideração os requisitos da curadoria digital nas diversas disciplinas científicas e acadêmicas e demais contextos de uso, compreendendo e explicando a curadoria digital para cada caso (CONSTANTOPOULOS; DALLAS, 2008).

Tendo em vista tais estratégias, Constantopoulos e Dallas (2008) se embasam no estudo anterior de Dallas (2007) para a formulação da abordagem. Nesse estudo, Dallas (2007) sugere que o escopo do trabalho de curadoria digital precisa de expansão, na medida em que o objetivo é assegurar a adequação epistêmica da informação em diversos contextos de uso. Para isso, a curadoria digital deve levar em conta que a validade e a utilidade dos recursos informacionais digitais dependem de uma representação adequada do conhecimento (DALLAS, 2007). Tal perspectiva leva em consideração que a agregação de valor e conhecimento consiste em um processo crucial para a curadoria digital.

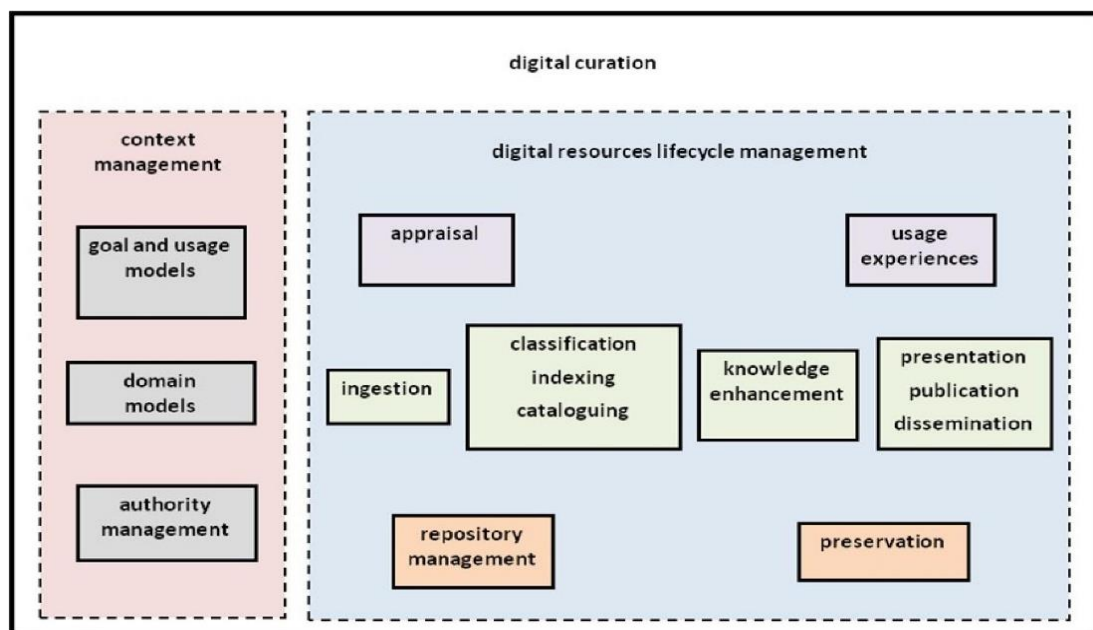
Tradicionalmente, para o universo dos patrimônios culturais, há um amplo espectro de informações que podem ser adicionadas aos recursos informacionais que possibilitam agregar valor informacional e conhecimento. Essas informações envolvem desde representações de objetos culturais reais, abrangendo suas descrições, relacionamentos, teorias, interpretações e opiniões, até as pessoas e organizações envolvidas na produção e operação tais objetos. (CONSTANTOPOULOS; DALLAS, 2008). Essas informações formam o corpo de conhecimento que envolve os patrimônios culturais, que por sua vez, caracteriza-se pela “[...] diversidade disciplinar, complexidade representacional e heterogeneidade, orientação

histórica, viés textual e caráter cumulativo.” (CONSTANTOPOULOS; DALLAS, 2008, p. 3, tradução nossa).

Desse modo, a abordagem do DCU para a curadoria digital está direcionada à necessidade de apoiar uma representação baseada nesse corpo de conhecimento que envolve e caracteriza o universo dos patrimônios culturais. Em outras palavras, a curadoria digital precisava abranger além dos recursos digitais, também as informações relativas ao contexto semântico no qual esses recursos estão inseridos ou relacionados, envolvendo o que os autores Constantopoulos e Dallas (2008) denominam como conteúdo epistêmico de objetos de informação. Na dimensão semântica, é importante levar em consideração que existem diversas perspectivas que podem ser atribuídas aos recursos informacionais. Tais perspectivas tem origem nas diferentes tradições disciplinares e podem trazer conceitos e terminologias que se sobrepõem, muitas vezes causando inconsistências. Além disso, há outra questão relacionada a dimensão semântica que diz respeito à possibilidade de criação de novos significados a partir das informações mediadas e coproduzidas pela experiência do visitante e do usuário (CONSTANTOPOULOS; DALLAS, 2008).

Logo, os aspectos semânticos são considerados uma camada adicional e de grande importância para agregação de valor no contexto da curadoria digital. Com esse viés, o modelo do DCU prevê uma abordagem que conduz a curadoria digital em dois grupos de processos, tal como apresentado na figura 6.

Figura 6 - Processos da curadoria digital do DCU



Fonte: Constantopoulos e Dallas *et al.* (2009).

O primeiro grupo de processos apresentados na figura 6 é denominado Gerenciamento do contexto (*Context management*). Nesse grupo, encontram-se três processos que dão suporte aos demais processos (CONSTANTOPOULOS; DALLAS, 2008; CONSTANTOPOULOS; DALLAS *et al.*, 2009):

- Modelagem de meta e de uso (*Goal and usage modelling*): dois processos que fornecem representações, sendo que a modelagem de meta se destina a capturar as intenções dos criadores e dos usuários de uma determinada classe de recursos digitais e a modelagem de uso visa capturar os padrões de uso dos recursos.
- Modelagem de domínio (*Domain modelling*): processo que produz ou refina as representações de conhecimento especializado sobre um domínio de interesse, na forma de ontologias e modelos conceituais.
- Gerenciamento de autoridade (*Authority management*): processo que lida com os vocabulários controlados usados por convenção para denotar conceitos, propriedades e relações ou suas instâncias, conhecidas como autoridades.

O outro grupo de processos é denominado como Gerenciamento do ciclo de vida dos recursos digitais (*Digital resources lifecycle management*). Nesse grupo, encontram-se oito processos (CONSTANTOPOULOS; DALLAS, 2008; CONSTANTOPOULOS; DALLAS *et al.*, 2009):

- Avaliação (*Appraisal*): estabelecimento de critérios para a avaliação e seleção de recursos com potencial para incorporarem os processos de curadoria subsequentes.
- Admissão (*Ingesting*): processo que incorpora os recursos selecionados à curadoria digital. Essa etapa envolve a criação de recursos digitais, a digitalização de recursos analógicos e a importação de recursos digitais de outras fontes.
- Classificação, indexação e catalogação (*Classification, indexing and cataloguing*): processos necessários para a produção e incorporação de índices lógicos destinados ao gerenciamento de informações. Isto é, consiste no processo de descrição e representação dos recursos digitais para identificação e interpretação acerca do seu conteúdo e contexto. Tais informações podem ser consideradas como recursos digitais secundários que podem ser produzidos durante o processo ou podem ser recursos incorporados.
- Aprimoramento do conhecimento (*Knowledge enhancement*): processo que se destina a agregar valor para enriquecer os recursos digitais com base em informações do contexto e do domínio de conhecimento a que pertence ou está

relacionado. Por exemplo, um recurso digital pode ser associado com entidades de uma ontologia ou vincular o recurso a outros recursos que se relacionam, ou então representar formalmente as situações ou eventos relacionados aos recursos, de modo a explicitar o inter-relacionamento entre eles.

- Apresentação, publicação e disseminação (*Presentation, publication and dissemination*): processo que envolve a geração de novos artefatos (científicos, acadêmicos, artísticos, etc.) a partir de recursos digitais primários ou secundários existentes.
- Experiência do usuário (*User experience*): este processo registra a interação entre usuários e recursos digitais, bem como os efeitos dessa interação, com o intuito de adequar a forma de disponibilização dos recursos aos usuários finais e para a disposição dos recursos no próprio sistema.
- Gerenciamento de repositório (*Repository management*): esse processo engloba o gerenciamento dos repositórios (centralizados ou distribuídos) que serão utilizados para armazenamento dos recursos, assim como trata dos mecanismos de acesso aos recursos, com definições de trocas, armazenamento e backups, incluindo também o gerenciamento dos motores de busca.
- Preservação (*Preservation*): processo que visa salvaguardar os recursos digitais com vistas a preservação à longo prazo, evitando os riscos decorrentes de causas físicas ou da obsolescência tecnológica. Envolve políticas de preservação e o uso de metadados apropriados.

Desse modo, a abordagem orientada a processos do DCU dá ênfase a necessidade de registrar e manter como as informações, que são armazenadas, curadas e preservadas durante o ciclo de vida, são utilizadas e acessadas pelos usuários (CONSTANTOPOULOS; DALLAS, 2009). Por isso, percebe-se que a grande preocupação em adequar as informações de contexto e domínio para o compartilhamento e reutilização por meio de instrumentos que permitam a organização e a representação do conhecimento, como as ontologias e modelos conceituais.

Seguindo essa abordagem do DCU, Constantopoulos *et al.* (2009) propuseram um conjunto de aprimoramentos ao modelo de ciclo de vida do DCC, que denominaram como uma proposta de extensão para o modelo do DCC. O DCC&U é o resultado da análise e ampliação do modelo de curadoria digital proposto pelo DCC, que foi sugerido usando como base alguns dos elementos-chave da abordagem orientada a processos do DCU. Assim, o

DCC&U leva em consideração aspectos que não foram considerados no modelo base. Esses aspectos se referem a três propostas que incidem sobre o ciclo de vida de curadoria digital.

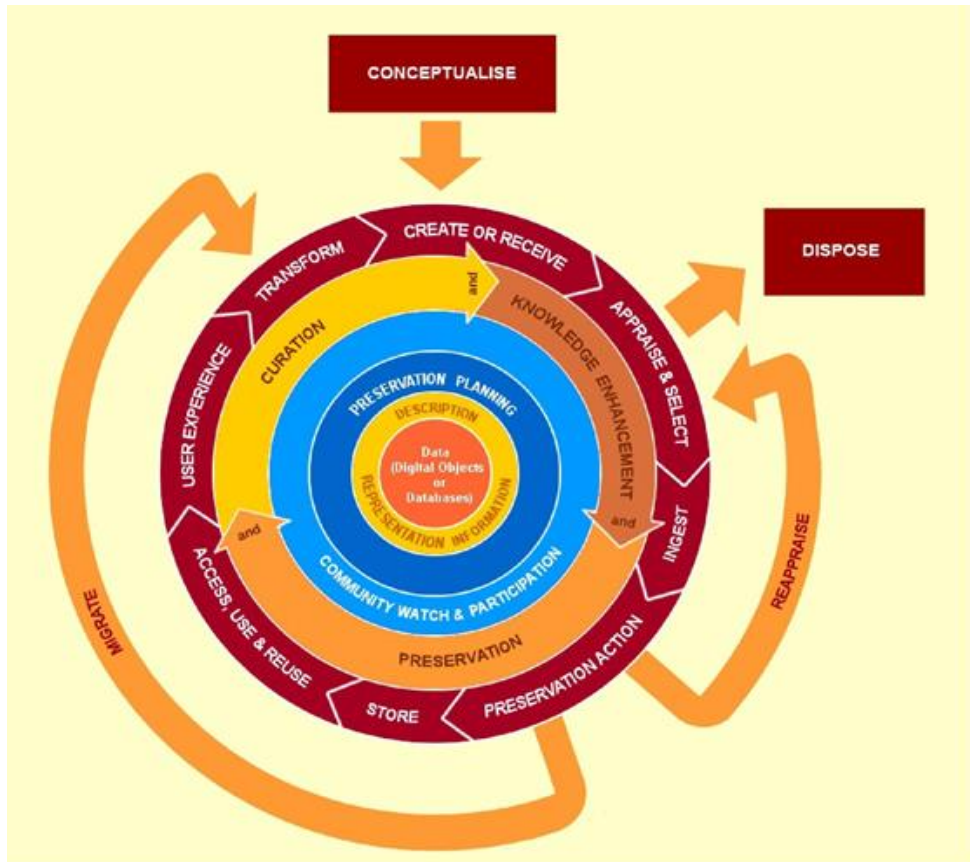
A primeira proposta leva em consideração que os usuários estão cada vez mais ativos, criando e modificando recursos digitais, formando comunidades e interagindo entre si. Assim, a proposta é adicionar uma ação nova após a ação sequencial Acesso e Reutilização, com o intuito de registrar e manter a experiência do usuário.

A segunda proposta se refere a inclusão da manutenção de autoridades, levando em consideração como elas evoluem com o tempo. Considerando que “[...] boa parte do conhecimento especializado sobre o domínio é capturado em autoridades usadas por convenção para denotar conceitos, propriedades e relações relevantes, ou suas instâncias.” (CONSTANTOPOULO *et al.*, 2009, p.41, tradução nossa) a segunda proposta sugere a ampliação da ação Descrição e Representação da Informação do modelo do DCC para incluir essas informações.

A terceira e última proposta de ampliação enfatiza o aprimoramento do conhecimento como um processo fundamental para a curadoria digital. Isto é, o conhecimento produzido tanto pela pesquisa acadêmica e científica, como pela prática profissional, geram constantemente novos conhecimentos que incidem sobre as entidades, situações e eventos do mundo real representados pelos recursos digitais, o que por sua vez, incide sobre o contexto e o domínio relacionados com os recursos digitais (CONSTANTOPOULOS; DALLAS *et al.*, 2009). Por essa razão, é proposta a ampliação da Ações para Todo o Ciclo de Vida existente no modelo DCC, de *Curate and Preserve* (Curadoria e Preservação) para *Curation, Preservation and Knowledge Enhancement* (Curadoria, Preservação, Aprimoramento do Conhecimento). De acordo com Constantopoulos e Dallas *et al.* (2009, p.42, tradução nossa), o aprimoramento do conhecimento é uma ação que precisa ser executada durante todo o ciclo de vida dos objetos digitais e configura-se como “[...] um componente essencial do modelo de ciclo de vida de curadoria, no qual o conhecimento é adicionado em cima de um repositório existente de recursos digitais e em toda a sua base de conhecimento relacionada.”

Assim, do mesmo modo que o DCC fornece uma visão geral gráfica de alto nível dos estágios necessários para empreender a curadoria digital, o modelo estendido é apresentado de forma gráfica. Na figura 7, é possível visualizar as três propostas de ampliação do modelo do DCC que ilustram a abordagem alternativa que foi sugerida para visualizar os processos abrangidos pela curadoria digital.

Figura 7 - Extended Digital Curation Lifecycle Model (DCC&U)



Fonte: Constantopoulos *et al.* (2009).

Por ser baseada em um dos modelos de curadoria digital mais discutido na literatura, tal abordagem se torna mais consistente e compreensível. No entanto, com os próprios autores alertam, essa é uma abordagem gradual focada na harmonização das ações e processos pontuados pelos modelos do Centro de Curadoria Digital (DCC) e da Unidade de Curadoria Digital (DCU) e que por isso não está completa. Os processos do Gerenciamento do contexto: Modelagem de metas e uso, Modelagem de domínio e Gerenciamento de autoridade, não são introduzidos no modelo DCC&U.

Apesar disso, o modelo DCC&U configura-se como um avanço para a teoria que envolve a curadoria digital, por defender que agregar valor aos recursos digitais abrange além da sua descrição propriamente dita, também a representação dos aspectos semânticos que envolvem esses recursos como uma camada adicional na curadoria digital, que permitirá assegurar a sua validade, utilidade, ou seja, a autenticidade, confiabilidade, integridade e usabilidade dos dados para adequação ao propósito informacional em diversos usos a longo prazo.

3.2 Análise e comparação dos modelos de curadoria digital

Diante dos aspectos pontuados pelos modelos, é possível destacar que todos eles tem em comum a importante função de estabelecerem diretrizes para a prática da curadoria digital em função da necessidade de se empreender planejamento, manutenção e interdependência de etapas. Percebe-se que desde as primeiras abordagens direcionadas ao gerenciamento do recursos digitais, destaca-se a necessidade do encadeamento entre o ciclo de vida do recurso digital e o meio em que este se encontra.

No modelo *Open Archival Information Systems (OAIS)*, o foco está em orientar a implementação de sistema de arquivamento digital, levando em consideração as atividades internas do sistema para a preservação digital e permitindo gerenciar o fluxo de informações.

Já o modelo do *Joint Information Systems Committee (JISC)* foi desenvolvido tendo em vista objetivos mais amplos, compreendo que o gerenciamento dos recursos digitais precisava levar em consideração além do sistema interno, também os aspectos dos recursos digitais e dos serviços distribuídos de hospedagem e acesso.

Considerando com mais ênfase esse relacionamento entre os aspectos dos recursos digitais e dos sistemas, o modelo do *Digital Curation Centre (DCC)* apresenta uma abordagem que se preocupada em contemplar todas as instâncias que se relacionam a manutenção dos recursos digitais, dando ênfase para a capacidade de garantir a sua integridade e autenticidade intelectual, bem como a confiabilidade, o uso e o reuso dos dados.

Esse fato fica ainda mais evidente na proposta do *Extended Digital Curation Lifecycle Model (DCC&U)*, que introduz a ideia de informação contextual com o objetivo de destacar a necessidade de atribuir informações que ajudem no gerenciamento dos recursos digitais de modo a potencializar sua capacidade de reutilização em diferentes contextos.

Assim, percebe-se que enquanto os dois primeiros modelos estão mais voltados para o sistema que suporta a curadoria digital, os dois últimos abordam a prática de uma perspectiva do objeto da curadoria digital, os dados.

No quadro 7, é sintetizado as etapas que cada um dos modelos estabelece, de modo apresentar uma comparação entre cada uma delas.

Nos modelos OAIS e JISC, as ações são mais sucintas, muitas delas vinculadas a outras ações. Nos outros dois modelos, do DCC e do DCC&U, é possível perceber um maior detalhamento ao indicar cada uma das ações, havendo um maior número delas. Esse maior detalhamento que pode ser percebido nos modelos do DCC e do DCC&U, ilustram a preocupação que os estudos em curadoria passaram a ter ao buscar a implementação de suas

abordagens. A complexidade do emprego de métodos para a gestão de dados fez com que os modelos se expandissem e considerassem novas necessidades.

Quadro 7 - Comparação entre os modelos de ciclo de vida de Curadoria

OAIS	JISC	DCC	DDC&U
Informação descritiva		Descrição e representação da informação	Descrição e representação da informação (Informações de autoridades)
Planejamento de preservação	Planejamento de preservação	Planejamento de preservação	Planejamento de preservação
		Acompanhamento e participação da comunidade	Acompanhamento e participação da comunidade
Administração do Sistema; Gerenciamento de dados	Armazenamento seguro e remoto	Curadoria e preservação	Curadoria, preservação e aprimoramento do conhecimento
		Conceptualização	Conceptualização
Recepção (Pacote de Informações de Submissão - PSI)		Criação ou recepção	Criação ou recepção
		Avaliação e seleção	Avaliação e seleção
	Admissão	Admissão	Admissão
Planejamento de preservação; Gerenciamento de dados		Ação de preservação	Ação de preservação
Armazenamento (Pacotes de Informação de Arquivamento - PAI)	Armazenamento	Armazenamento	Armazenamento
Acesso (Pacote de Informação de Divulgação - PDI)	Acesso	Acesso, uso e reuso	Acesso, uso e reuso
			Experiência do usuário
		Transformação	Transformação
		Descarte	Descarte
		Reavaliação	Reavaliação
		Migração	Migração

Fonte: Elaborado pela autora.

Contudo, a visualização do quadro 7 permite inferir que existem ações que são recorrentes a todos os modelos, apesar de apresentarem nomenclaturas diferentes. Essas ações podem ser consideradas um denominador comum entre os modelos de ciclo de vida de

curadoria digital apresentados. Verifica-se que as ações de Armazenamento, Acesso e Planejamento da preservação aparecem nos quatro modelos. Outra ação que apresenta correspondência é a de gerenciamento que, no modelo OAIS correspondem às ações de Administração do Sistema e Gerenciamento de dados, no modelo JISC, à ação de Armazenamento seguro e remoto, no do DCC, à ação de Curadoria e preservação e no DCC&U a ação de Curadoria, preservação e aprimoramento do conhecimento.

Verifica-se, assim, que todos os modelos levam em consideração, cada um a seu modo, que os dados precisam existir de tal forma que possam transmitir uma informação significativa. Para tanto, na relação apresentada no quadro 7, constatou-se que as ações que envolvem a representação da informação são fundamentais para os modelos, não constando explicitamente apenas no modelo do JISC, que elenca somente os processos de armazenamento no local; replicação para repositórios de arquivamento de terceiros confiáveis e externos; e o planejamento de preservação. Porém, na execução de cada um desses processos a representação está implícita.

Portanto, reforça-se a ideia de que o gerenciamento de recursos informacionais está diretamente relacionado com sua representação, envolvendo os múltiplos aspectos que influenciam o ciclo de vida dos dados nos ambientes digitais.

Esta análise das similaridades e as diferenças entre os modelos forneceu um panorama das ações que caracterizam a curadoria digital, proporcionando a base para elucidar a discussão proposta, que será apresentada a seguir.

4 RELAÇÕES ENTRE METADADOS E CURADORIA DIGITAL

Diante dos aspectos que caracterizam a curadoria digital, pode-se vislumbrar que para a Ciência da Informação, ela se configura como uma possível perspectiva das formas de tratamento contemplada pela organização, representação e manutenção da informação em meio digital.

Para a Ciência da Informação, o processo de representação da informação se destaca como um campo fundamental do tratamento informacional, estando relacionado ao estabelecimento de identificação, de ordem e de registro unívoco de um recurso informacional visando sua recuperação, acesso e localização (ALVES, 2010).

Na curadoria digital, o tratamento informacional ocorre a partir da contínua gestão dos dados, visando reduzir as ameaças ao valor informacional e também o risco de obsolescência, se valendo para isso da representação da informação.

A Representação da Informação, como subcampo da Ciência da Informação tem sido dividida para fins didáticos em representação descritiva, a qual lida com os aspectos específicos para a individualização, identificação e preservação do recurso informacional, e a representação temática, que está voltada para os aspectos de assunto, ou seja, para a temática que engloba o recurso. Ambas as representações são fundamentais para possibilitar uma eficiente recuperação de informações e ocorrem na prática de modo integrado e colaborativo (MAIMONE; SILVEIRA; TÁLAMO, 2011). Principalmente no âmbito digital, ambas tendem a ser executadas de modo associado, em uma única plataforma, por meio dos metadados.

Para a curadoria digital, os metadados são parte fundamental de sua atividade, pois estabelecem relação direta com os dados, que estão no centro de suas ações, caracterizando os chamados objetos digitais (HIGGINS, 2008). Pode-se, então, destacar a sinergia da curadoria digital com a Ciência da Informação, por ter como preocupação central a representação dos recursos informacional. Como explica Araújo (2017), a curadoria digital nasce como um campo prático e profissional, que tem, no entanto, desdobramentos ou consequências teóricas, vista a preocupação em estabelecer princípios norteadores para suas ações. São esses princípios que passam a ter grande interesse para a Ciência da Informação, sendo sua principal contribuição a “[...] preocupação com o todo, isto é, com a ligação e a interdependência entre os vários aspectos, momentos e instâncias relacionados com a informação” (ARAÚJO, 2017, p.15).

Por conseguinte, a curadoria digital tem suas ações apoiadas nos metadados para gerir os dados; isto é, são os metadados que viabilizam a descrição e o controle adequados dos dados a longo prazo, de tal modo a garantir a sua autenticidade, confiabilidade, integridade e usabilidade (HIGGINS, 2008). Verifica-se que há tempos a questão da manutenção da autenticidade, da confiabilidade, da integridade e da usabilidade das informações é considerada por setores tradicionais como as instituições LAM, no entanto, ela toma proporções ainda maiores, diante do crescente volume de informações e do constante desenvolvimento das tecnologias no âmbito digital.

Como destaca Harvey (2010, p. 80, tradução nossa):

Os bibliotecários e arquivistas estão muito familiarizados com a necessidade de metadados para apoiar atividades de curadoria, mas para curadoria de objetos digitais os requisitos são mais extensos, especialmente aqueles para metadados de preservação. Para os indivíduos que criam objetos digitais, os metadados são igualmente importantes.

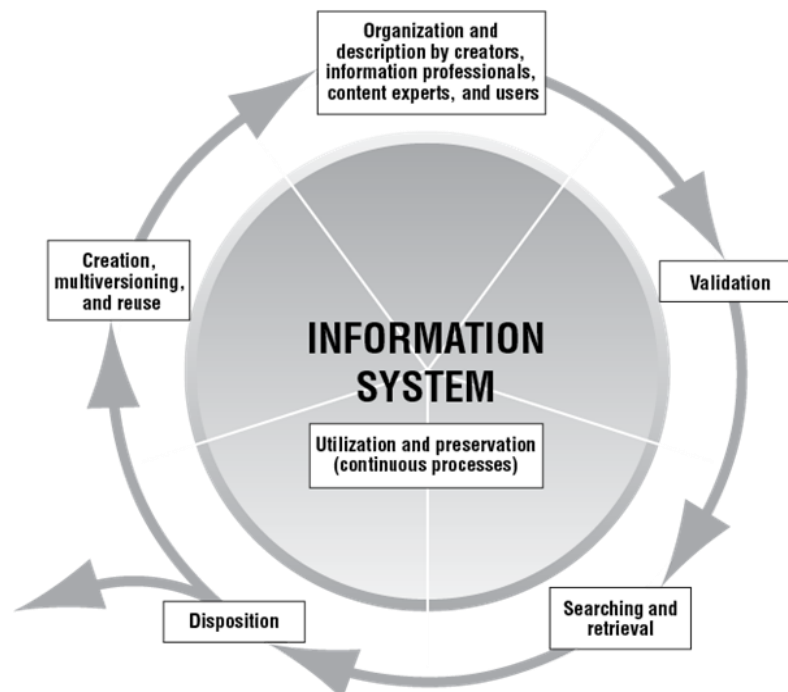
A necessidade de metadados para apoiar atividades de curadoria não é algo novo para os profissionais da informação, como bibliotecários, museólogos e arquivistas. Contudo, para curadoria digital, uma ampla gama de aspectos deve ser considerada. Como é possível ser visualizado na figura 8⁶, Gilliland (2016), ilustra a complexidade de processos a que os metadados estão associados por meio das diferentes fases pelas quais os recursos informacionais se movem durante seus ciclos de vida no ambiente digital.

Como um modelo em formato cíclico que representa a vida de um recurso informacional ao longo do tempo, as fases são contínuas e interdependentes. Os recursos iniciam sua trajetória a partir da fase de “*Creation, Multiversioning and Reuse*” (Criação, Multiversão e Reutilização), que corresponde a fase em que são inseridos no ciclo, sejam estes natos digitais ou digitalizados, novos ou reutilizados ou ainda provenientes de um processo de transformação. Em sequência, está a fase de “*Organization and description by creators, information professional, content experts and Users*” (Organização e descrição por criadores, profissionais da informação, especialistas do conteúdo e usuários), que diz respeito ao processo de tratamento informacional do recurso de tal forma que ele seja corretamente identificado e processado. Nessa fase, há a participação de diversos profissionais, desde os criadores até os usuários finais, para que os recursos sejam devidamente representados atendendo ao seu propósito informacional. A fase seguinte corresponde a ação de

⁶ Elaborada por Gilliland (2016), com base na figura “*Information Life Cycle*” elaborada por Borgman, C. L. *et al.*, no relatório “*Social Aspects of Digital Libraries* da *UCLA-NSF Social Aspects of Digital Libraries Workshop, Los Angeles, CA: Graduate School of Education and Information Studies* 1996, p. 7.

“*Validation*” (Validação), na qual os recursos passam por um processo de teste para verificar a efetividade do tratamento informacional. Em seguida os recursos passam para a fase de “*Searching and retrieval*” (Pesquisa e recuperação), processo em que eles são pesquisados e recuperados pelos seus usuários. Após isso, os recursos informacionais passam para a fase de “*Disposition*” (Disposição), na qual os recursos que não possuem mais uma função no sistema informacional são descartados, e os que ainda possuem, reintegram a trajetória, começando no ciclo de vida.

Figura 8 - Ciclo de vida dos recursos informacionais no ambiente digital



Fonte: Gilliland (2016).

Em cada uma dessas fases, os recursos informacionais adquirem vários tipos de metadados ao percorrer o ciclo, criando-se uma mistura complexa de processos e camadas. De acordo com Gililand (2016), essas camadas de metadados podem ser associadas de várias maneiras. Uma forma de associar metadados nesse processo é embuti-los no próprio arquivo dos recursos informacionais, tal como em geral ocorre com os arquivos de imagem, ou então a partir de um esquema que indique os tipos de metadados, bem como os relacionamentos entre as várias partes do recurso. Outra forma de associar metadados está relacionada com a vinculação desses elementos a partir de *hyperlinks* em conjunto com a criação de registros de

metadados que determinem os relacionamentos entre os metadados e os recursos e entre os próprios metadados.

Como visto na seção três, sendo a abordagem da curadoria digital baseada no ciclo de vida dos recursos informacionais, pode-se vislumbrar que a mesma complexidade relacionando suas ações aos metadados é enfrentada, em especial para no contexto de gerenciamento e preservação dos recursos digitais. O quadro 8, elaborado com base em Harvey (2010), sintetiza os tipos e as funções dos metadados e seus aspectos que podem ser considerados para a curadoria digital.

Quadro 8 - Metadados e suas funções para a curadoria digital

Função geral	Tipo	Função específica	Exemplos
Descrever dados e sua localização	Metadados Descritivos	Identificar os dados para que possam ser vinculados a solicitações	Nome do criador do conjunto de dados; Nome do autor de um documento.
	Metadados Estruturais	Descrever como os objetos digitais compostos são organizados	Relação da imagem da página TIFF (<i>Tagged Image File Format</i>) com outras imagens da página.
Fornecer as informações técnicas necessárias para usar dados	Metadados Técnicos	Fornecer as informações técnicas necessárias para usar dados	Formato; Algoritmos de compressão ou codificação; Chaves de criptografia e descryptografia de Software (incluindo o número da versão) usados para criar ou atualizar os dados.
		Fornecer informações sobre o ambiente geral do sistema	Hardware, sistemas operacionais, software aplicativo nos quais os dados foram criados.
Descrever o que acontece com os dados conforme eles se movem no ciclo de vida da curadoria	Metadados Administrativos	Fornecer informações sobre os processos de uso, gerenciamento e codificação de objetos digitais durante um período de tempo	Informações sobre a criação de dados, atualizações subsequentes, transformação, controle de versão, resumo; Descrições de migração e replicação.
	Metadados Preservação	Registrar as ações de preservação que foram aplicadas aos dados ao longo do tempo	Formato de arquivo; Propriedades significativas; Ambiente técnico; Informações de correção.

Fonte: Adaptado de Harvey (2010).

Diante disso, é possível elucidar que os metadados são fundamentais para a curadoria digital. No entanto, concorda-se com Harvey (2010) que o grande desafio é fornecer quantidades adequadas de metadados e que eles sejam apropriados não só para o processamento por humanos, mas também pelas máquinas. Por essa razão, destaca-se a importância do uso de padrões e de melhores práticas desenvolvidos pelas comunidades

interessadas, a fim de garantir qualidade, consistência e interoperabilidade, também de modo automatizado.

Por essa razão, um dos grandes desafios das atividades orientadas a dados é a necessidade de garantir a descoberta, o acesso, a análise, o uso e o reuso de dados de modo apropriado para demandas e para fluxos de trabalho, concomitantemente para humanos e para máquinas.

Tendo em vista esses desafios, principalmente no âmbito da gestão e da administração de dados científicos, um conjunto de princípios com a finalidade de tornar os dados localizáveis, acessíveis, interoperáveis e reutilizáveis, foi criado. Denominados Princípios FAIR, (acrônimo para: *Findability, Accessibility, Interoperability e Reusability*), o termo foi lançado em 2014, tendo sua aplicação consolidada em 2017, diante da exigência, por parte da Comissão Europeia, de que os projetos financiados por seus recursos deveriam adotar um plano de gestão de dados com base nesses princípios (HENNING *et al.*, 2018).

Desse modo, os Princípios FAIR enfatizam

[...] a capacidade de processamento da máquina (isto é, a capacidade dos sistemas computacionais de localizar, acessar, interoperar e reutilizar dados sem intervenção humana mínima), porque os humanos dependem cada vez mais do suporte computacional para lidar com os dados como resultado do aumento no volume, complexidade e velocidade de criação de dados (PRINCÍPIOS FAIR, [2019], não paginado, tradução nossa).

Para que os sistemas computacionais sejam capazes de realizar tais tarefas, é necessário que os dados sejam devidamente representados de tal modo a suportar esse alto nível de automação. Por isso, os Princípios FAIR são considerados diretrizes para o “bom” gerenciamento de dados, incluindo também os metadados (BISHOP; HANK, 2017).

O primeiro princípio refere-se à capacidade de localizar os dados. Dados e metadados precisam ser passíveis de serem encontrados, sendo legíveis tanto para os humanos quanto para os sistemas automatizados. O segundo princípio diz respeito à capacidade de acessar os dados, o que pode incluir processos de autenticação e autorização. O terceiro princípio leva em consideração que os dados não são usados de modo isolado, sendo em geral integrados a outros dados. Por isso, esse princípio se refere à capacidade de interoperar, sejam entre os dados, os metadados ou com os aplicativos para análise, armazenamento e processamento. O quarto e último princípio faz referência ao objetivo final do FAIR, que consiste em otimizar a reutilização de dados a partir de dados e metadados de boa qualidade para que estes possam ser replicados e combinados em diferentes configurações e contextos (FAIR PRINCIPLES, [2019]). O quadro 9 sintetiza os Princípios FAIR.

Quadro 9 - Princípios de Dados FAIR

Localizável (<i>Findable</i>)	F1. Os (meta)dados recebem um identificador globalmente exclusivo e persistente
	F2. Os dados são descritos com metadados ricos (definidos por R1 abaixo)
	F3. Os metadados incluem clara e explicitamente o identificador dos dados que descrevem
	F4. Os (meta)dados são registrados ou indexados em um recurso pesquisável
Accessível (<i>Accessible</i>)	A1. Os (meta)dados são recuperáveis pelo identificador usando um protocolo de comunicação padronizado A1.1 O protocolo é aberto, gratuito e universalmente implementável A1.2 O protocolo permite um procedimento de autenticação e autorização, quando necessário
	A2. Os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis
Interoperável (<i>Interoperable</i>)	I1. Os (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
	I2. Os (meta)dados usam vocabulários que seguem os princípios do FAIR
	I3. Os (meta)dados incluem referências qualificadas a outros (meta)dados
Reutilizável (<i>Reusable</i>)	R1 (meta)dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes R1.1 (Meta)dados são liberados com uma licença de uso de dados clara e acessível R1.2 (Meta)dados estão associados com proveniência detalhada R1.3 Os (meta)dados atendem aos padrões da comunidade relevantes ao domínio

Fonte: adaptado de Princípios FAIR, [2019].

Como é possível visualizar no quadro 9, as diretrizes fazem referência aos dados e aos metadados de modo acoplado, como ‘(meta)dados’. Isso permite inferir que os princípios se aplicam tanto aos dados, quanto aos próprios metadados.

Da mesma forma, para a curadoria digital, tais diretrizes devem ser consideradas para um efetivo gerenciamento dos dados, levando em consideração também os metadados, ou seja, que as informações geradas e atribuídas para os objetos digitais sejam curadas conjuntamente no ciclo de vida (HARVEY, 2010). Por isso, é possível encontrar pontos de correspondência entre os princípios pontuados em FAIR e os preceitos que orientam a curadoria digital.

Segundo Princípios FAIR ([2019], não paginado, tradução nossa),

Os princípios se referem a três tipos de entidades: dados (ou qualquer objeto digital), metadados (informações sobre esse objeto digital) e

infraestrutura. Por exemplo, o princípio F4 define que os metadados e os dados são registrados ou indexados em um recurso pesquisável (o componente de infraestrutura).

Na curadoria digital, ocorre algo semelhante, levando em conta os aspectos centrados no usuário (ser humano), no sistema (máquina) e no conteúdo (recurso informacional). Por isso, uma importante consideração a ser feita sobre a relação entre a curadoria digital e os metadados é o fato de que este último possibilita agregar valor informacional aos recursos informacionais.

Segundo Beagrie (2006), a curadoria digital é uma atividade que, por se beneficiar do uso já existente da “curadoria” em setores tradicionais, como bibliotecas, museus e setores das ciências biológicas, leva em consideração além da preservação e da manutenção de uma coleção ou banco de dados, também algum grau de valor informacional agregado e conhecimento.

Tradicionalmente, a atribuição de valor informacional às coleções tem estado centrada na figura do curador. Principalmente em bibliotecas e museus, a curadoria está alinhada à criação de coleções temáticas em torno de objetos físicos, bem como de sua documentação, estando fortemente relacionada ao conhecimento de domínio e a atribuição de contexto, de modo que este seja relevante para pesquisa, descoberta e aprendizado (BEAGRIE, 2006).

Paralelamente, um dos principais objetivos da curadoria digital é garantir o valor informacional dos recursos digitais para os propósitos a que foram destinados. Como salientam Lee e Tibbo (2011, p.127, tradução nossa), “[...] em contraste com o cuidado de materiais analógicos, o curador digital se depara com uma ampla gama de oportunidades e desafios”, principalmente quando a organização e a representação do conhecimento de domínio e a atribuição de contexto são vinculados a processos automatizados, dadas as enormes quantidades de dados envolvidos.

Tendo em vista que os recursos em ambientes digitais, principalmente na *Web* de Dados, tendem a ser utilizados para diferentes propósitos, a curadoria digital precisa de mecanismos que vão além das ações de manutenção, implicando em agregar valor para que esses recursos possam atender aos diferentes propósitos informacionais a que são destinados. Por essa razão, a curadoria digital tende a se beneficiar ao estar alinhada aos preceitos da *Web* Semântica e de suas tecnologias de modo a continuar fornecendo abordagens e estratégias que garantam o gerenciamento dos dados, com qualidade e confiabilidade, mesmo em ambientes abertos e pouco controlados.

Percebe-se, então, que o fato da curadoria estar apoiada nos metadados implica também que ela é sustentada pelas novas dimensões que envolvem o tratamento da informação em diferentes ambientes digitais. Na Ciência da Informação, tais dimensões relacionam-se aos campos da Organização e Representação da Informação que historicamente estão relacionados com a aplicação de tecnologias (ALVES, 2010; ARAÚJO, 2014). Segundo Araújo (2014), os estudos voltados para a representação da informação vêm passando por grandes mudanças, principalmente devido às tecnologias em meio digital e as novas possibilidades no que diz respeito à descrição, à classificação e à organização da informação.

Em tal cenário, as tecnologias semânticas ganham destaque, influenciando diretamente os aspectos da organização e da representação da informação e do conhecimento, e conseqüentemente, da curadoria digital. Para tanto, é possível encontrar uma grande variedade de padrões de metadados, vocabulários controlados e ontologias orientados para enriquecer a representação dos recursos informacionais a partir de soluções semânticas, de modo a garantir um maior nível de interação, de integração e de precisão na recuperação da informação.

Nota-se que, a partir da ampliação do uso das tecnologias semânticas, diversas abordagens tem surgido para dar conta dos desafios relacionados ao tratamento de dados na *Web*, configurando um novo e promissor cenário de colaboração e de integração propícios para a curadoria digital. O estudo desse cenário permite verificar como os metadados possibilitam atribuir e assegurar o contexto semântico dos recursos informacionais no processo de curadoria digital.

4.1 A contextualização semântica na curadoria digital

Em uma retrospectiva pela evolução da curadoria digital, percebe-se que ela surge tentando se diferenciar da curadoria tradicional, buscando criar novas abordagens orientadas às possibilidades e aos desafios que os ambientes digitais passaram oferecer.

No entanto, a análise da literatura permite interpretar que na medida em que as tecnologias evoluem e que as máquinas passam a exercer em grande parte as atividades humanas, principalmente no que diz respeito às formas de produção, à análise e ao uso de grandes quantidades de dados e de informações, a curadoria digital não se afasta conceitualmente de suas raízes, permanecendo voltada para garantir o propósito informacional dos recursos por meio da contextualização semântica.

Em Linguística, semântica é o termo atribuído ao estudo do significado e da interpretação do significado de um signo, ou seja, de uma palavra, sentença ou expressão em

um determinado contexto. Na Ciência da Informação, a semântica está relacionada com o significado das relações entre os recursos informacionais (GLUSHKO, 2013), constituindo o que Constantopoulos e Dallas (2008) reconhecem como o corpo de conhecimento que envolve e forma o domínio dos patrimônios culturais. Tal corpo de conhecimento é constituído por todas as informações que representam um recurso informacional, abrangendo suas descrições, relacionamentos, teorias, interpretações, opiniões, bem como os atores envolvidos em sua produção e operação.

Em setores tradicionais, como museus, galeria de artes e bibliotecas, a curadoria pode ser comumente compreendida de duas formas: como a prática de organizar mostras e exposições de uma área específica; ou, então, como um conjunto de técnicas que visam à conservação de objetos físicos, compreendendo todo o ciclo de atividades relativas aos acervos, desde os trabalhos acerca da formação e desenvolvimento de coleções; estudo, documentação, comunicação, acesso, divulgação, circulação, até eventuais procedimentos de manutenção e restauro, para fins de formação profissional, científicos e educacionais (SANJAD; BRANDÃO, 2008).

Em ambos os casos, todo esse corpo de conhecimento que caracteriza o universo dos patrimônios culturais, ou seja, o seu contexto semântico, configura-se como a base para orientar as ações que serão realizadas sobre o acervo ou coleção, seja para a criação do ambiente em que uma exposição ou mostra é planejada e montada, ou para definição dos critérios de seleção e de preservação a serem considerados em um ambiente informacional.

Independentemente do modo como tais ações de curadoria são realizadas, tradicionalmente a atribuição de contexto semântico às coleções e aos objetos analógicos está centrada na figura de um humano especialista de domínio, o curador. Para Silva e Loureiro (2015), o curador é um especialista em determinada temática e uma figura decisória que detém o conhecimento e a capacidade de estabelecer seleção e ordem para a apresentação de um determinado conteúdo.

Em contrapartida, na curadoria digital, observa-se que a contextualização semântica não é realizada apenas por especialistas humanos, mas por um conjunto de tecnologias, princípios e profissionais que fazem com que as máquinas possam exercer em grande parte esse papel.

Por esse motivo, constata-se que a grande diferença entre o tradicional processo de curadoria e a curadoria digital é a existência de uma interface não humana entre o conhecimento de domínio e o usuário. Essa interface, que tradicionalmente era desempenhada apenas por humanos, passa então a ser feita também por processos automatizados,

responsáveis em responder às demandas dos usuários a partir da codificação dos conceitos e das relações entre os dados.

Como implicação, o conhecimento de domínio precisa ser organizado e representado também para as máquinas. Para tanto, há necessidade de instrumentos que considerem a dimensão semântica dos dados para o processamento computacional.

Desde o início da automação dos processos de representação da informação no âmbito das instituições LAM, identifica-se a preocupação em criar estruturas que considerem os aspectos sintáticos e semânticos dos recursos informacionais para o processamento computacional. Questões relacionadas a representação de recursos informacionais no ambiente digital estão em debate desde a década de 1960 (ASSUMPÇÃO; SANTOS, 2015), período anterior ao desenvolvimento da *Web*, culminando no desenvolvimento do formato MARC (*Machine Readable Cataloging*), e posteriormente, dos formatos MARC21, amplamente utilizado em bibliotecas para promover a comunicação de registros bibliográficos com informações padronizadas legíveis por computador. Sua estrutura leva em consideração campos e subcampos que correspondem à sintaxe do registro bibliográfico, sendo estes preenchidos de acordo com regras de conteúdo, ou seja, definições semânticas para padronizar a descrição bibliográfica.

Além dos formatos MARC, demais estruturas foram impulsionadas pelos avanços das Tecnologias da Informação e Comunicação (TIC) resultando em diversos instrumentos, princípios, modelos e tecnologias relacionadas à representação da informação no âmbito computacional.

Dentre essas estruturas, as denominadas tecnologias semânticas passaram a desempenhar um papel determinante na comunicação de informações padronizadas legíveis por computador, influenciando de diversas formas no modo de organização e de representação da informação e do conhecimento.

De acordo com Ramalho e Ouchi, (2011, p.66) “[...] as tecnologias semânticas caracterizam-se como linguagens que possibilitam ir além de representações sintáticas, descrevendo computacionalmente aspectos semânticos dos documentos [...]” Trabalhando de modo integrado para que os recursos sejam bem representados pelos metadados, as tecnologias semânticas atuam em diferentes aspectos da representação da informação objetivando que as máquinas processem os dados de modo semanticamente coerente com o contexto.

Retomando os estudos da *Digital Curation Unit* (DCU), o denominado Gerenciamento do contexto (*Context management*), proposto pelos pesquisadores do centro

(CONSTANTOPOULOS; DALLAS, 2008; CONSTANTOPOULOS; DALLAS *et al.*, 2009), dá ênfase na necessidade de adequar as informações de contexto para a curadoria digital por meio de estruturas que permitam a organização e a representação do conhecimento para as máquinas, como modelos conceituais e ontologias.

No âmbito da representação da informação, as ontologias assumem a função de explicitar os modelos conceituais, permitindo representar um conjunto de conceitos e de relacionamentos dentro de um domínio (RAMALHO, 2006; FUSCO, 2010).

As ontologias são consideradas as estruturas centrais das tecnologias semânticas para explicitar os elementos (metadados) que caracterizam um recurso informacional, visando a integração, comunicação e troca de informações estruturadas. Para tanto, Doerr (2008) discute dois dos usos mais importantes das ontologias para o universo da curadoria digital.

Em primeiro lugar, o autor destaca a ontologia tal como uma linguagem, que dá suporte à indexação ao explicitar as possíveis relações semânticas entre os recursos informacionais, garantindo a descrição dos ‘significados pretendidos’. Doerr (2008) aponta que “[...] uma ontologia conhecida pelo indexador e pelo usuário pode ser considerada como substituto do diálogo verbal tradicional entre bibliotecários e usuários de bibliotecas”. Do mesmo modo, é possível fazer uma correspondência em relação à atividade curatorial em meio digital, em que a comunicação entre o curador e o público passa a ser feita por meio da codificação de um sistema.

Campos (2009, p. 10), explica que “[...] as relações semânticas com propósito de representação do conhecimento, ao serem explicitadas, eliminam interpretações dúbias: dois conceitos podem se relacionar de muitas formas diferentes, mas ao explicitar a relação pretendida, o conhecimento é comunicado.” Isso se torna particularmente importante para a curadoria digital por permitir apoiar seus processos a partir da extração de conhecimento, da integração de conhecimento e do raciocínio automatizado (DOERR, 2008).

Tal raciocínio automatizado é o que Harvey (2010) aponta como solução para as limitações dos recursos humanos na curadoria digital, considerando as enormes quantidades de dados envolvidos; ou seja, as ontologias podem influenciar na cadeia de processos que compõe a atividade curatorial, em especial no contexto da *Web*, em ambientes que lidam com volumosos conjuntos de dados heterogêneos e com processos distribuídos.

Outra importante função das ontologias na curadoria digital apontada por Doerr (2008) está relacionada às atividades no nível do sistema, como *design*, configuração, integração de esquemas de metadados. As ontologias permitem que *designers* de sistemas, curadores digitais e especialistas em conteúdo possam estar de acordo com os requisitos funcionais dos

sistemas, definir estruturas de dados e metadados e identificar mecanismos para obter interoperabilidade e integração entre sistemas heterogêneos.

Como visto na seção três, a curadoria digital considera todos os estágios do ciclo de vida de um recurso informacional, incluindo suas relações com outros recursos. Para isso leva em consideração um alto nível de planejamento, de manutenção e de interdependência de etapas, o que por sua vez, exige a comunicação entre suas ações. Portanto, a explicitação dos significados pretendidos permite otimizar a comunicação entre as ações da curadoria digital para orientar o fluxo de dados, potencializando o raciocínio automatizado.

Hellman (2009, não paginado, tradução nossa), compara que “[...] enquanto os metadados das bibliotecas tradicionais sempre estiveram direcionados a ajudar os seres humanos a encontrar e usar informações, as ontologias da *Web* estão focadas em ajudar as máquinas a encontrar e usar informações”, tornando os metadados significativos e processáveis para as máquinas. Para tanto, as ontologias ganham espaço no universo da organização e representação da informação e do conhecimento e podem ser consideradas grandes aliadas da curadoria digital, pois possibilitam

[...] que as máquinas processem as informações fornecidas pelos agentes humanos de maneira consistente com o significado pretendido. Como tal, elas desempenham um papel importante na mediação entre usuários e sistemas de informação, mas também entre especialistas em domínio e especialistas em TI, e na orientação e justificação de decisões de design de sistemas (DOERR, 2008, p.28, tradução nossa).

Evidencia-se que o uso de ontologias permite apreender o domínio em que a curadoria digital está sendo empregada, abrangendo as diferentes instâncias do gerenciamento do ciclo de vida e o contexto semântico a que um conjunto de dados que está sendo curado pertence. As ontologias representam formalmente os elementos de corpo do conhecimento, o que permite adequar as informações a cada finalidade e ao seu propósito informacional para cada contexto de uso.

Além disso, a curadoria digital leva em consideração a integração de diversos tipos de dados para poder atender aos requisitos de gerenciamento e manutenção em seu ciclo de ações, o que reflete diretamente no uso dos diversos tipos de metadados, e conseqüentemente em diferentes esquemas e padrões de metadados. A integração de metadados estruturados de maneiras diferentes, no entanto, requer algum recurso que permita explicitar os significados dos dados em cada uma das estruturas para que eles possam ser corretamente codificados.

Sendo assim, as ontologias tem grande importância para a curadoria digital, na medida em que viabilizam a interoperabilidade entre esquemas heterogêneos, permitindo a

comunicação semântica e sintática entre os elementos de metadados. Para isso, é importante que a interoperabilidade entre várias ontologias também seja observada, permitindo assim que diferentes comunidades reutilizem e compreendam de modo apropriado as informações ao longo do tempo (DOERR, 2008).

Infere-se que a interoperabilidade é um dos alicerces para o compartilhamento e reuso dos dados. De acordo com Harvey (2010, p. 68, tradução nossa), “[...] o compartilhamento e a reutilização de dados, que estão no centro da curadoria digital, exigem interoperabilidade, o que, por sua vez, exige aderência aos padrões”. Como já visto na seção dois, para o sucesso da proposta da *Web Semântica* é necessário que a estruturação dos dados seja aderente a padrões amplamente compartilhados (POMERANTZ, 2015). Portanto, o uso de recomendações e também princípios configuram importantes requisitos a serem levados em consideração para a curadoria digital.

Nesse aspecto, ocupando a proposta central da *Web Semântica* está o denominado *Linked Data*. Como já discutido na seção dois, o *Linked Data* refere-se à possibilidade de conectar dados relacionados que não foram previamente ligados (BIZER; HEATH; BERNERS-LEE (2009), possibilitando a reutilização de dados já publicados em diferentes contextos e em sistemas heterogêneos.

Para a curadoria digital, o emprego dos princípios do *Linked Data*, bem como do *Open Data* (dados com licença aberta) reflete-se na melhoria da qualidade da representação informacional, fornecendo a possibilidade de conectar e atribuir dados e metadados apropriados a partir dos relacionamentos semânticos entre os recursos.

Como discute Marcondes (2018), tanto o processo de publicação, quanto o de representação da informação, usando tecnologia *Linked Open Data* (LOD), permite criar o que o autor denomina como ‘*links* semânticos’, que são os possíveis relacionamentos culturalmente relevantes que os recursos do patrimônio cultural estabelecem entre si, a partir da vinculação de seus dados.

Frequentemente essas coleções apresentam relacionamentos culturalmente relevantes entre seus objetos, como por exemplo, um livro sobre uma pintura, um esboço ou esboço de uma pintura famosa, uma carta de um autor comentando um livro ou uma obra de arte, ou um contrato para encomendar uma escultura ou obra de arte etc. (MARCONDES, 2018, p. 1076).

Esses relacionamentos culturalmente significativos delineiam o contexto semântico a que cada recurso do patrimônio cultural pertence, dando a possibilidade de vincular diversos acervos e coleções a partir das suas complementariedades, agregando valor a esses recursos informacionais. É importante considerar, ainda, que a estrutura mínima de descrição é o que

garante que os dados sejam passíveis de serem vinculados e disponibilizados para diversos usos.

Logo, com o intuito de alcançar além do uso e preservação momentâneos, a curadoria digital busca agregar valor por meio do contexto e da ligação entre os dados, dando ênfase para a criação e a publicação de dados de maneira a facilitar a reutilização e promover integridade e a integração (RUSBRIDGE *et al.*, 2005).

O processo de agregar valor na curadoria digital, portanto, se destina enriquecer os dados com base em informações do contexto semântico, ou seja, metadados definidos a partir do corpo de conhecimento a que um recurso informacional pertence ou está relacionado. Entende-se que o processo de agregar valor informacional na curadoria digital está relacionado a assegurar a validade e utilidade dos recursos, isto é, a autenticidade, a confiabilidade, a integridade e a usabilidade dos dados para adequação a qualquer que seja seu propósito informacional ao longo do tempo.

De acordo com o Dicionário Brasileiro da Língua Portuguesa Michaelis (online), esses termos podem ser definidos da seguinte forma: Autenticidade é a natureza, propriedade ou condição do que é autêntico e legítimo; Confiabilidade é a qualidade de algo que é confiável, fiável; Integridade é relativo ao estado ou característica de algo que está inteiro, completo; Usabilidade é a capacidade de facilitar e otimizar o uso de algo (MICHAELIS, 2020). Todas essas características são complementares e se unem para dar caráter de qualidade aos dados.

A partir dessas definições e tendo como base o modelo do DCC é possível estabelecer a relação entre essas características e as ações pertencentes à curadoria digital.

A autenticidade e a confiabilidade podem ser garantidas pelos processos que certificam a proveniência dos dados, como na Ação Sequencial de '*Avaliação e seleção*', que vai analisar se os dados são aderentes às orientações documentadas, bem como às políticas e às exigências legais. Outras Ações Sequenciais que vão influenciar nos aspectos de autenticidade e a confiabilidade são as ações de '*Admissão*' e '*Armazenamento*', que levam em consideração formatos e estruturas de dados aderentes a padrões.

Já a integridade, que tem direta influência no grau de autenticidade e a confiabilidade de um recurso - portanto também sendo influenciada pelas ações de '*Avaliação e seleção*', '*Admissão*' e '*Armazenamento*' - será impactada principalmente pela Ação Sequencial de '*Conceptualização*', na qual ocorre o planejamento de quais dados farão parte da curadoria digital, e pela Ação Sequencial de '*Criação e/ou Recepção*', na qual os dados que foram julgados necessários para o gerenciamento serão criados ou recebido de fontes externas (agregadores, repositórios ou *data centers*), levando em consideração os tipos de metadados.

Por isso, a Ação Sequencial de ‘*Transformação*’, também exercerá influência na integridade, pois ela tem a finalidade de criar novos dados a partir dos originais, seja para adequá-los a uma determinada aplicação ou para mantê-los sustentáveis ao longo do tempo de modo a contemplar as necessidades de gerenciamento e manutenção.

Em relação à usabilidade, a Ação Sequencial que mais pode afetar essa característica é principalmente a de ‘*Acesso, uso e reuso*’, que considera a qualidade dos dados para viabilizar acesso, tanto pela publicação de forma aberta, levando em conta direitos e licenças, como a partir do controle de acesso por procedimentos de autenticação.

Percebe-se assim, que devido à complementariedade desses conceitos e da própria interdependência das ações da curadoria digital, mais de uma ação pode influenciar cada característica. O que fica bastante claro devido ao fato de que as quatro características são influenciadas pela ‘*Ação de preservação*’, que tem como intuito garantir que os dados permaneçam autênticos, confiáveis e utilizáveis, mantendo a respectiva integridade. Para isso, tal ação inclui procedimentos de limpeza de dados, validação, atribuição de metadados padronizados e de estruturas ou formatos de dados correspondentes.

Nesse sentido, reitera-se que a importância dos metadados para a curadoria digital pode ser percebida não só como um fator essencial para garantir o planejamento e a execução de suas ações, mas principalmente para garantir que os recursos digitais tenham seu propósito informacional assegurado independente do contexto de uso, garantindo sua qualidade e utilidade para o acesso, o uso e o reuso a longo prazo.

Sendo assim, concorda-se com Sayão (2016, p.49) ao esclarecer que o arcabouço prático e conceitual oferecido pela curadoria digital:

[...] permite a elaboração de fluxos de trabalho voltados para a uma gestão dinâmica de coleções de materiais digitais que podem ser aplicados aos acervos culturais digitais, ampliando o seu potencial de reuso e, como desdobramento, a concepção e o desenvolvimento de serviços *online* inovadores e de espaços de interação em torno dos acervos digitais.

É o que acontece no caso da *Europeana*, que a partir da iniciativa da comunidade Europeia de reunir e estruturar dados de diversas instituições como bibliotecas, arquivos, museus e galerias de arte, tornou-se uma grande plataforma interativa que não só disponibiliza o patrimônio cultural das instituições europeias, como também, está relacionada ao *OpenGLAM* (ARAKAKI; SIMIONATO; SANTOS, 2017), um movimento de instituições para promover o uso e o acesso aberto de dados, oferecendo um detalhado programa de contextualização semântica para agregar e reutilizar dados.

Tal como a *Europeana*, demais iniciativas surgem com o apoio do governo, como é o caso da *Digital Public Library of America* (DPLA) que busca reunir e estruturar o conteúdo sobre o patrimônio cultural dos Estados Unidos (EUA), com uma proposta bastante semelhante à da *Europeana*, e a ainda a *DigitalNZ*, criada a partir da proposta de conectar o conteúdo digital da Nova Zelândia e implementada pela a Biblioteca Nacional e pelo governo do país.

Operando cada uma a seu modo, as três plataformas tem em comum a proposta de integrar, estruturar e disseminar dados de patrimônios culturais na *Web* por meio instrumentos de representação da informação apoiados nas tecnologias semânticas. De acordo com Liu, Bikakis e Vlachidis (2017, p. 344, tradução nossa), no âmbito dessas plataformas, “[...] as tecnologias da *Web Semântica* são usadas principalmente para duas finalidades: o desenvolvimento de sistemas de curadoria interna e o estabelecimento de bancos de dados abertos de coleta.” Em ambos os casos, as tecnologias semânticas permitem a padronização dos dados que descrevem as entidades relativas aos patrimônios culturais e seus relacionamentos, possibilitando, dessa forma, a interoperabilidade e a sustentabilidade para diferentes coleções de dados.

O desenvolvimento da *Europeana* se enquadra nas atuais tendências de organização e de representação em ambientes digitais que levam em consideração os aspectos semânticos e a ligação entre os dados para empreender a curadoria digital de patrimônios culturais.

4.2 A curadoria digital na plataforma *Europeana*

A *Europeana*, também chamada de Biblioteca Digital *Europeana*, configura-se como uma plataforma interativa que reúne e estrutura dados de diversas instituições que trabalham com patrimônio cultural - como bibliotecas, arquivos, museus, entre outras - com o objetivo de disponibilizar amplamente o conteúdo sobre o patrimônio cultural da Europa, promovendo o reuso e o acesso aberto aos dados.

Seu projeto teve início em 2005, motivado pelo ideal de tonar o vasto conteúdo sobre patrimônios culturais europeus mais acessíveis. Sua proposta norteadora foi de “[...] criar um espaço no qual todas as manifestações da herança cultural e científica da Europa pudessem ser conectadas e integradas em um único portal, em um ambiente multilíngue.” (PURDAY, 2009, p. 920, tradução nossa). Com forte apoio político, a proposta era criar um ambiente digital que pudesse integrar vários domínios e disponibilizar o vasto conhecimento europeu.

Como explicam Concordia, Gradmann e Siebinga (2010), o coração do projeto foi um esforço para construir uma plataforma aberta que promovesse a interoperabilidade funcional, técnica e de dados. Nesse sentido, a interoperabilidade entre domínios foi um princípio fundamental a ser levado em conta desde o início do projeto, partindo da “[...] ideia de que a colaboração e a transferência de conhecimento entre os museus, bibliotecas, arquivos e coleções audiovisuais era fundamental para o sucesso do empreendimento.” (PURDAY, 2009, p. 921, tradução nossa). Sendo assim, no decorrer do desenvolvimento do projeto, as questões relacionadas à interoperabilidade dos dados foram cuidadosamente estudadas, levando em consideração padrões, princípios, vocabulários e requisitos de usuários.

Após alguns anos de testes e aprimoramentos, com foco no processo de aquisição de conteúdo e preparação de dados, a *Europeana* tornou-se em 2010 um sistema operacional. No mesmo ano houve a criação de um comitê denominado de “Comitê dos Sábios”, que tinha o objetivo de criar recomendações referentes à disponibilização, à digitalização, ao acesso e à preservação do conteúdo. Com isso, a *Europeana* ganhou força e novos apoiadores, chegando a ter, em novembro de 2010, um acervo de mais de 25 milhões de registros (WINER; ROCHA, 2013). Em 2019, a plataforma contava com mais de 58 milhões de itens, como livros, áudios, vídeos, obras de arte e coleções temáticas especiais que incluem arte, moda, música, fotografia e a documentos da Primeira Guerra Mundial.

Um importante marco na história da *Europeana* ocorreu no final de 2012, quando, de modo inédito, a plataforma disponibilizou sob a licença *Creative Commons* CC0 1.0, seus metadados de forma gratuita para vários tipos de uso (EUROPEANA PRO, 2017). Isso representou um grande salto em direção aos ideais da plataforma, permitindo que “[...] todos os metadados (informações textuais sobre patrimônio cultural digitalizado) no *site* [...]” fossem publicados sem quaisquer restrições de reutilização (EUROPEANA COLLECTIONS, [2020], não paginado, tradução nossa). Desde então, em constante aprimoramento, a *Europeana* ainda mantém o foco na interoperabilidade entre os domínios que lidam com o patrimônio cultural, dedicando grande parte dos seus esforços para manter um sistema funcional que garanta o livre acesso a um conteúdo de qualidade.

Além disso, seu grande potencial está na possibilidade de disseminar a terceiros seu conteúdo por meio de *Application Programming Interface* (API) específico, permitindo amplo reuso de seus dados estruturados (MARCONDES, 2016). Um API consiste em um conjunto de funções, que podem ser usadas para interagir com um programa de computador, possibilitando que um usuário exporte dados ou importe dados de um serviço da *Web* (POMERANTZ, 2015). Essa tecnologia propicia, portanto, que a *Europeana* ofereça um

serviço na *Web* de acesso remoto às suas coleções e ainda, que diversas comunidades façam uso de suas estruturas de representação de dados.

A construção da *Europeana* se alicerça na adoção de tecnologias e no desenvolvimento de padrões. Para a estruturação de seus dados, a *Europeana* desenvolveu uma ferramenta para mapear e enriquecer metadados denominado como *Europeana Data Model* (EDM) (EUROPEANA, 2013). O EDM configura-se como um modelo de dados fundamentado nas tecnologias da *Web Semântica* e nos princípios do *Linked Open Data*.

Desenvolvido a partir de um esforço colaborativo entre os setores do patrimônio cultural, o EDM mantém esse esforço colaborativo para suportar extensões e refinamentos, a fim de acomodar as particularidades dos dados do domínio do patrimônio cultural (CHARLES; ISAAC, 2015). Para isso, a *Europeana* busca manter o EDM como um modelo flexível e com vasta documentação para apoiar o trabalho específico de cada comunidade.

Como pioneira no projeto de agregar dados de diversas fontes em um ambiente informacional aberto como a *Web*, a *Europeana* enfrentou diversos desafios para colocar sua plataforma em funcionamento e ainda hoje trabalha para maximizar o potencial das suas fontes de informação com o intuito de transformar dados em algo mais significativo que possa ser amplamente usado e explorado pelos usuários. A *Europeana* demonstra grande empenho em ampliar seu conteúdo de alta qualidade e aprimorar sua infraestrutura compartilhada, melhorando a reutilização e permitindo enriquecer os dados para todas as partes interessadas.

Para isso, a *Europeana* considera uma estrutura de três níveis para o gerenciamento de sua plataforma, sendo o primeiro nível chamado de Núcleo, no qual coletam-se os dados e reúne-se conteúdo e tecnologia; um segundo nível, que é o Nível de Acesso, no qual se padroniza e enriquece-se o núcleo, definindo as regras de engajamento e fornecendo as interfaces de acesso; e por fim, um terceiro nível, que é denominado Nível de Serviço, no qual se desenvolvem experiências personalizadas para três de comunidades de usuários: profissionais, finais e criativos (desenvolvedores e empreendedores).

No Núcleo é onde são armazenados e gerenciados os recursos informacionais e seus metadados com tecnologia de código aberto. Nesse nível há a preocupação em usar tecnologia emergente para garantir o melhor cenário possível para o armazenamento dos recursos digitais em formas mais baratas e acessíveis, mantendo os *links* entre os dados sempre persistentes de modo que as informações sejam legíveis por humanos e máquinas. Além disso, há uma grande preocupação em continuar a investir em princípios como *Linked Open Data* para tornar os dados ainda mais funcionais (EUROPEANA STRATEGY 2015-2020, 2017).

O segundo nível da estrutura da plataforma lida com o acesso ao material, dando condições e diretrizes para acessá-lo. Nessa camada são estabelecidas as regras de engajamento entre os parceiros provedores de dados e a plataforma, levando em consideração as restrições de direitos autorais, as políticas e os modelos de negócios. Também é considerado o formato em que os dados estão estruturados e se são suficientes para permitir ampla distribuição de modo aberto e vinculável (EUROPEANA STRATEGY 2015-2020, 2017).

Por último, há o terceiro nível, que diz respeito aos serviços da plataforma para atender as demandas dos usuários. Nesse nível, parte-se do princípio de que em um meio aberto e distribuído como a *Web*, as demandas por informação ainda não estão totalmente estabelecidas, nem mesmo o grau de especialização dos usuários. Para tanto, prioriza-se trazer à tona o potencial do conteúdo, criando canais de interesse especializados para cada setor, bem como dando condições, melhores ferramentas e redução de custos para o compartilhamento de conhecimento entre parceiros contribuintes e usuários finais, buscando desenvolver conteúdo de qualidade tanto amplo, como também especializado (EUROPEANA STRATEGY 2015-2020, 2017).

Em relação às comunidades, três delas são atendidas. Uma é a comunidade dos parceiros contribuintes, que formam o grupo de profissionais composto pelas instituições que digitalizam e compartilham as coleções, provendo os dados para a *Europeana*. A outra é a dos usuários finais, que correspondem a todas as pessoas que utilizam ou que são potenciais utilizadores do conteúdo oferecido pela plataforma. E, finalmente, na terceira comunidade, encontram-se os chamados usuários criativos, que são pessoas ou grupos de desenvolvedores e empreendedores que criam novos produtos e serviços. Esses são conhecidos coletivamente como "indústrias criativas" e demandam altos níveis de serviço e acesso a um conteúdo reutilizável e de qualidade (EUROPEANA STRATEGY 2015-2020, 2017).

Constata-se, diante do apresentado, que a *Europeana* realiza a curadoria digital de seus dados. A partir do estudo da documentação disponibilizada pela plataforma, foi possível identificar diversos pontos de convergência entre as ações que caracterizam a curadoria digital, apresentadas na seção três, e o gerenciamento proposto pela *Europeana*, que tem por base o *Europeana Data Model* (EDM).

Na parte central das ações de gerenciamento da plataforma, encontram-se os registros formados pelos pacotes de dados sobre um recurso do patrimônio cultural. Tais pacotes de dados compreendem as informações que relacionam os metadados, representações digitais dos recursos, e demais dados contextuais, como os agentes, locais, conceitos, etc. Esses registros

são agrupados por conjuntos de dados, que são considerados como uma coleção de dados ou como um pacote de informações conforme definido pelo Modelo OAIS (EUROPEANA PRO, [2020]).

Um conjunto de dados pode ser enviado diretamente para a *Europeana* a partir de um fornecedor, como as intuições do patrimônio cultural, ou então por um provedor de dados, ou seja, o conjunto de dados é fornecido a partir de um agregador de conteúdo externo. Esses dados podem ser a respeito de um determinado assunto, e ser proveniente de uma fonte específica ou uma coleção mantida por uma instituição; ou, ainda, um conjunto de dados pode ser agregado sem correspondência direta com uma coleção específica (EUROPEANA PRO, [2020]).

Independente da fonte, todos os registros recebem um identificador exclusivo ao serem incorporados à *Europeana*. Para isso, o EDM utiliza o *Internationalized Resource Identifier* (IRI), que consiste em uma extensão do esquema *Uniform Resource Identifier* (URI), permitindo identificar todos os dados, incluindo os metadados de um recurso digital, bem como, o próprio recurso para que eles possam ser referenciados sem ambiguidade.

Destacam-se, dessa forma, os requisitos do EDM que regem o fluxo de entrada dos dados:

- R1: distinção entre "objetos fornecidos" (pintura, livro, filme, sítio de arqueologia, arquivo de arquivo, etc.) e suas representações digitais;
- R2: distinção entre objetos e registros de metadados que descrevem um objeto;
- R3: vários registros para o mesmo objeto devem ser permitidos, contendo declarações potencialmente contraditórias sobre esse objeto;
- R4: suporte para objetos que são compostos de outros objetos;
- R5: compatibilidade com diferentes níveis de abstração de descrição (por exemplo, se um fornecedor deseja enviar descrições que seguem as distinções introduzidas no Grupo 1 do FRBR);
- R6: o EDM fornece um formato padrão de metadados que pode ser especializado;
- R7 suporte para recursos contextuais, incluindo conceitos de vocabulários controlados. (EUROPEANA, 2013, p. 8, tradução nossa).

Para que isso seja efetivo, os dados fornecidos à *Europeana* pelos provedores precisam atender aos critérios de publicação definidos pela plataforma, visando proporcionar melhor qualidade dos dados e uma maior possibilidade de conexão entre as coleções. Para tanto, o grupo *Europeana* desenvolveu o *Europeana Publishing Guide*, (disponível em: <https://pro.europeana.eu/post/publication-policy>), um conjunto de diretrizes que define os critérios para a apresentação de dados (EUROPEANA, 2019). Dentre os motivos destacadas para o uso do guia está a busca para garantir que os dados sejam sempre vistos como

autênticos, confiáveis e robustos pelo público (EUROPEANA, 2017b). Tal preocupação revela a importância das políticas para o gerenciamento de dados.

Após admitidos, atentando aos critérios e aos princípios estabelecidos, os conjuntos de dados são incorporados à plataforma. Com base no estudo de Haslhofer e Isaac (2011), Coneglian e Santarem Segundo (2017) descrevem o processo pelo qual os dados são submetidos ao serem incorporados à *Europeana* por meio do EDM.

Tal processo consiste em uma conversão que se inicia no momento da obtenção dos dados, no qual eles são transformados para o EDM. Essa conversão transforma os dados em recursos representados pelo padrão RDF/XML, fornecendo um identificador de recursos na *Web*. Em seguida, é realizado o processo de enriquecimento dos dados, no qual são inseridos mais dados relevantes sobre os objetos culturais pela própria *Europeana*. A partir desse estágio, os dados são ligados a demais recursos com URIs já existentes em outras bases, o que lhes confere maior representatividade (CONEGLIAN; SANTAREM SEGUNDO, 2017).

Nessa etapa, é onde ocorre o processo de agregar valor informacional aos conjuntos de dados para que eles não apenas sejam devidamente identificados e recuperados, mas principalmente para que possam fomentar ainda mais conexões dentro do seu contexto semântico. Nesse ponto, o gerenciamento de contexto proposto pelo *Digital Curation Unit* (DCU) no *Extended Digital Curation Lifecycle Model* (DCC&U) ganha um exemplo concreto.

Segundo a *Europeana* (2017a, p. 8, tradução nossa),

O modelo de dados da *Europeana* (EDM) tem como objetivo ser um meio de integração para coletar, conectar e enriquecer as descrições fornecidas pelos provedores de conteúdo da *Europeana*. Como tal, pode-se dizer que inclui qualquer elemento (ou seja, classe ou propriedade) encontrado na descrição de um provedor de conteúdo.

O EDM é capaz de mapear várias fontes e integrar diversos tipos de dados considerando as diferentes perspectivas e necessidades de informação dos usuários para uma melhor representação, permitindo assim, multidimensionar as formas de interação entre plataforma e cada comunidade usuária.

O modelo não se baseia em nenhum padrão comunitário específico, mas adota uma estrutura aberta baseada na *Web Semântica* que pode acomodar a variedade e a riqueza de padrões específicos de cada setor (EUROPEANA, 2013). Sendo assim, o EDM tem como objetivo geral explorar a riqueza dos dados disponíveis na *Web* sobre patrimônio cultural para uma representação mais completa, buscando apoiar a integração dos vários modelos que usam dados do patrimônio cultural, “[...] para que todas as descrições originais possam ser coletadas

e conectadas por meio de conceitos de nível superior.” (EUROPEANA, 2013, p. 8, tradução nossa).

Isso é notável nos três ‘princípios de *design*’ nos quais o modelo está fundamentado:

- D1: O EDM permite a integração de dados em um ambiente aberto: é impossível antecipar todos os dados contribuídos;
- D2: O EDM permite uma funcionalidade rica, possivelmente por meio de extensões;
- D3: O EDM deve reutilizar os modelos existentes (padrão) tanto quanto possível (EUROPEANA, 2013, p. 8, tradução nossa).

Para tanto, o EDM permite o enriquecimento de dados por meio de uma variedade de fontes externas. Como explicam Coneglian e Santarem Segundo (2017, p. 91), o EDM “[...] utiliza diversos vocabulários para formar um modelo que consiga expressar as ligações existentes, entre autores, obras, organizações, direitos autorais, além de outros tipos de informações, contidas em um objeto cultural.” O modelo trabalha, portanto, com uma diversidade de elementos que são fornecidos à plataforma, formando conjuntos de dados abertos e flexíveis que se expandem à medida que novos provedores passam a interagir com a plataforma (EUROPEANA, 2017a). Isso significa que o EDM é capaz de reunir e estruturar os dados de modo que várias descrições coexistam para um mesmo item fornecido.

Como explicam Haslhofer e Isaac (2011, p. 98, tradução nossa), a estrutura central de dados do EDM está:

[...] baseada no Modelo de Reutilização de Objetos de Arquivo Aberto (OAI-ORE), um modelo de referência para a descrição e troca de agregações de recursos da Web. As agregações de ORE são usadas para representar uma dada contribuição do provedor para a *Europeana*, que consiste no “item fornecido” junto com sua “visão(ões)” digital (modelada como recursos da Web).

Sendo assim, cada item fornecido à *Europeana* poderá vir com seus próprios metadados, sendo que o EDM será capaz de reuni-los e integrá-los para uma representação mais abrangente e não excludente. Por essa razão, o EDM é usado para fins de harmonização e integração de dados (HYVÖNEN, 2012), pois viabiliza que os diversos dados fornecidos sejam estruturados a partir de diferentes perspectivas e necessidades informacionais.

Doerr *et al.* (2010, p. 6-7, tradução nossa) explicam que o EDM “[...] não vincula a representação de metadados ingeridos a um esquema comum, mas age como uma ontologia de nível superior comum, de acordo com a qual os metadados compatíveis com outros modelos de dados originais podem ser expressos.” Sendo assim, o EDM pode ser considerado como uma âncora, que permite conectar vários modelos mais refinados, ao mesmo tempo em que garante a interoperabilidade em um nível semântico (CHARLES; ISAAC, 2015).

Um exemplo de modelo que está alinhados ao EDM, é o CIDOC-CRM (*CIDOC Conceptual Reference Model*), principal ferramenta para a modelagem de dados no campo do patrimônio cultural. Esse alinhamento funcional com demais modelos, faz com que o EDM suporte uma descrição mais completa do conteúdo ao mesmo tempo em que se preocupa com a definição de perfis de aplicação para que os dados sejam adequados a cada finalidade.

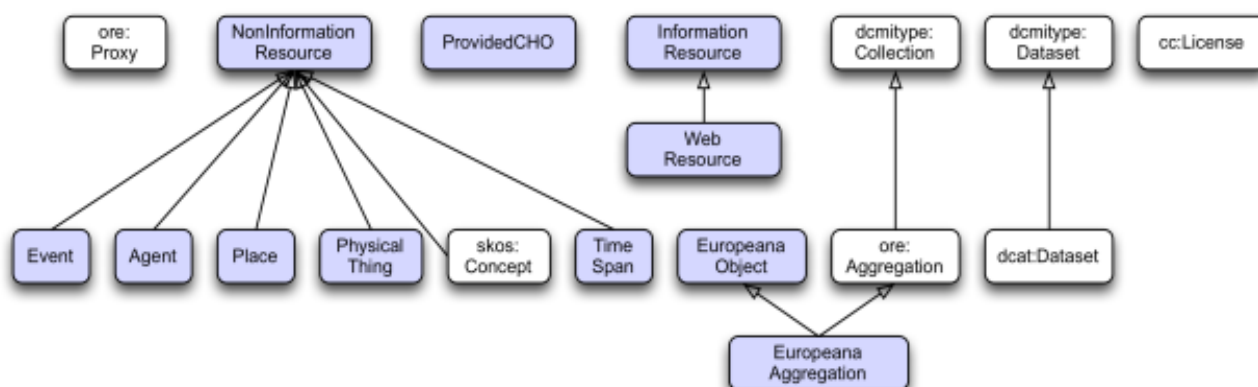
Um perfil de aplicação pode ser definido como:

Um conjunto de elementos de metadados, políticas e diretrizes definidas para um aplicativo específico. Os elementos podem ser de um ou mais conjuntos de elementos, permitindo que um determinado aplicativo atenda a seus requisitos funcionais usando metadados de vários conjuntos de elementos, incluindo conjuntos definidos localmente (DUBLIN CORE METADATA GLOSSARY, 2001, não paginado, tradução nossa).

A partir da abordagem por perfis de aplicação, o EDM faz mais do que mesclar e modificar os elementos existentes, ele cria estruturas modulares que permitem integrar e expandir vários vocabulários já existentes, acrescentando ainda seus próprios metadados. Seguindo o princípio da modularidade, o EDM reutiliza elementos de metadados externos e se alinha a outros modelos conceituais, possibilitando criar perfis de aplicação específicos e enriquecidos por diferentes provedores de dados (ZENG; QIN, 2016).

Para alcançar tal objetivo, o EDM está baseado em diretrizes e padrões reconhecidos internacionalmente, como *Open Archives Initiative Object Resue & Exchange* (OAI-ORE), *Dublin Core* (DC), *Simple Knowledge Organization System* (SKOS), *Resource Description Framework* (RDF) e *Uniform Resource Identifier* (URI), entre outros (HYVÖNEN, 2012; DOERR *et al.*, 2010; ARAKAKI, 2016). Tais tecnologias estabelecidas para a *Web Semântica*, discutidas na seção dois, permitem que o EDM atue como uma ontologia que reúne outros modelos de dados, ao mesmo tempo em que possibilita a interoperabilidade e a integração com os dados disponíveis na *Web*.

O EDM funciona a partir das declarações em RDF, definindo classes e propriedades em formato de triplas (sujeito, predicado, objeto). Essa definição se apresenta em hierarquia de classes, como ilustrado na figura 9, que demonstra como são definidos os relacionamentos entre os dados no EDM.

Figura 9 - Hierarquia de classes do EDM

Fonte: Europeana (2017a).

A figura 9 demonstra as relações semânticas que o EDM estabelece ao ligar-se a elementos já existentes, reutilizando vocabulários internacionalmente reconhecidos ao em vez de criar novos. As classes do EDM são representadas pelos retângulos azuis e as classes reutilizadas de outros padrões, pelos retângulos brancos.

É possível perceber que o modelo fornece estruturação para adaptar-se o trabalho específico de cada comunidade de patrimônios culturais. De acordo com Coneglian e Santarem Segundo (2017, p. 91) “O uso e integração de vocabulários internacionalmente reconhecidos e nascidos de padrões de metadados já sedimentados, torna o EDM um rico modelo de descrição dos objetos da *Europeana*, fortalecendo a estrutura de descrição semântica dos recursos.” Desse modo, a estrutura de representação do EDM também se torna bastante aceita pelas comunidades.

Para o processo de estruturar os dados, o EDM possui três classes principais de recursos que são provenientes do pacote de dados que deve ser fornecido à *Europeana*. São elas:

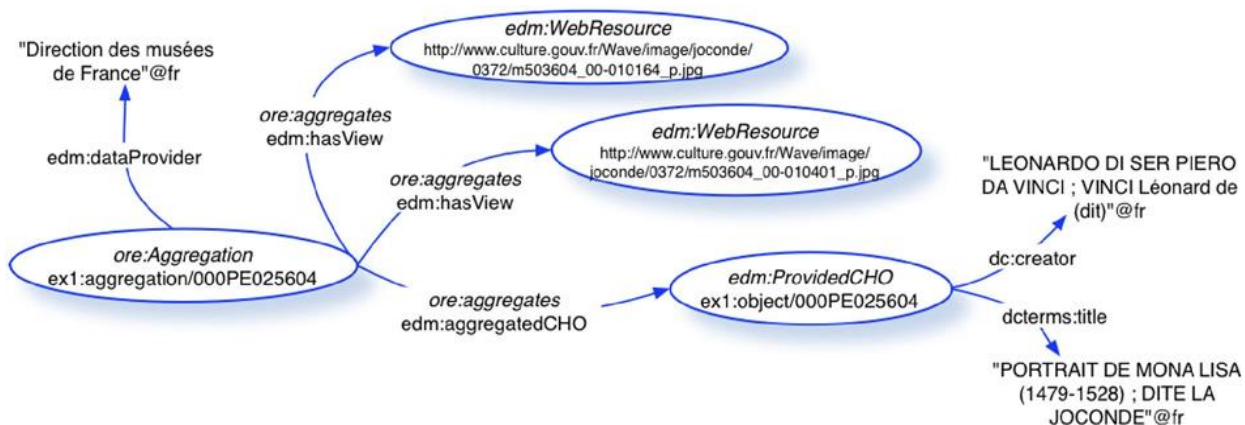
- o “objeto de patrimônio cultural fornecido” em si (uma pintura, um filme, uma partitura musical, um livro...) (edm: ProvidedCHO)
- uma ou mais representações digitais acessíveis deste objeto, algumas das quais serão usadas como visualizações (a imagem digital da pintura.) (edm: WebResource)
- uma agregação para representar o resultado da atividade desse provedor. (ore: Aggregation) (EUROPEANA, 2013, p.10, tradução nossa).

O elemento *ProvidedCHO*, refere-se ao objeto original que está sendo descrito, o elemento *WebResource* diz respeito às representações digitais dos objetos que estão na *Web* e

o elemento *Aggregation*, vincula os outros itens, fornecendo dados sobre a origem da descrição. Ainda, existem classes que podem ser agregadas para representar os recursos contextuais de cada elemento específico, tais como: *Agent*, *Place*, *TimeSpan*, *Concept*, *Event* e *PhysicalThing*. (EUROPEANA, 2013).

A atribuição desses elementos descritivos permite fornecer uma representação mais completa, que não exclui outras já existentes, ao contrário, agrega demais descrições para melhorar a representação do recurso de patrimônio cultural no seu domínio semântico. A figura 10 exemplifica como ocorre a representação em EDM.

Figura 10 - Representação em EDM da pintura Mona Lisa, de Leonardo da Vinci



Fonte: Europeana (2013).

Como é possível visualizar, a figura 10 apresenta um panorama da relação entre as principais classes e propriedades que as ligam. No caso, a obra *Mona Lisa*, de Leonardo da Vinci é representada pelo '*edm: ProvidedCHO*' e está relacionada a outras propriedades, como '*ore: Aggregation*', que indica os dados do seu provedor (*edm:dataProvider*), a '*Direction des Musées de france*'. Ainda, a obra é relacionada, por meio da atribuição de elementos descritivos como '*dc: criator*' e '*dc: title*', ao seu criador e ao seu título, ambos com valores atribuídos pela vinculação de dados aberto de outras bases de dados. Além disso, por meio das subpropriedades '*ore: agregates*', a obra é relacionada a outras representações digitais que já estão na *Web*, representadas pelo '*edm:WebResource*' (EUROPEANA, 2013).

Nesse contexto, destacam-se novamente os princípios do *Linked Open Data* (LOD), uma vez que a publicação de dados conectados de forma aberta viabiliza a integração de várias fontes, otimizando a representação dos dados e ampliando as possibilidades de colaboração.

No entanto, é importante destacar que as vantagens oferecidas pela veiculação aberta de dados de diversas fontes trazem ao mesmo tempo grandes desafios a serem levados em consideração, principalmente em relação à qualidade dos dados.

Tendo isso em vista, a comunidade de especialistas, desenvolvedores e pesquisadores da *Europeana*, denominada *EuropeanaTech*, reuniu-se em fevereiro de 2015 para uma conferência que trouxe como assunto principal o tema *Data Quality* (qualidade dos dados). Na ocasião foram discutidas as principais questões que afetam a qualidade de seus dados. Os questionamentos e as recomendações registrados no relatório da conferência permitem contribuir para compreender como ocorre a prática da curadoria digital no ecossistema da plataforma *Europeana*.

A questão central da conferência foi tentar entender por qual motivo os dados continuam com má qualidade se os provedores trabalham com estruturas bem definidas. Uma das conclusões a que todos os participantes da conferência concordaram foi a de que a maioria dos problemas relacionados à qualidade dos dados vem em decorrência do princípio de agregação. Mais precisamente, o problema decorre devido às muitas ações diferentes a que os dados são submetidos no fluxo de trabalho estabelecido pela *Europeana* para agregar dados. Essas diferentes ações podem resultar na melhoria da qualidade dos dados, mas por outro lado, podem levar à sua diminuição (EUROPEANA PRO, 2015).

Esse fluxo de trabalho envolve:

Desde a criação e entrega dos dados, até a definição do mapeamento, normalização e enriquecimento, os dados são definidos e redefinidos de acordo com diferentes padrões de dados. Também é preparado para diferentes ambientes e consumidores (EUROPEANA PRO, 2015, não paginado, tradução nossa).

Sendo assim, existe uma longa cadeia de processos para o fornecimento de dados, o que requer muita comunicação entre diferentes agentes, humanos e automatizados, devido às divergências e diferentes perspectivas acerca dos dados. Novamente, a comunicação entre as ações que compõe o gerenciamento dos dados é um fator essencial para a efetividade do fluxo de gerenciamento.

Além disso, outro problema apontado na conferência que afeta a qualidade dos dados é falta de informações sobre as ferramentas destinadas a apoiar o provisionamento de dados. Tais ferramentas destinam-se a realizar tarefas de exportação, mapeamento e transformação dos dados de uma estrutura para outra, sendo que sua falta ou uso ineficaz pode afetar a qualidade dos dados (EUROPEANA PRO, 2015). Ou seja, a deficiência na documentação também gera entraves para o gerenciamento dos dados.

Outra questão bastante importante levantada pelos participantes da conferência foi a respeito da efetividade do enriquecimento de dados como uma solução para obter dados de melhor qualidade. Quando um conjunto de dados fornecidos apresenta deficiências que afetam a semântica pretendida ou não estão de acordo com as necessidades especificadas, os dados que estão ausentes devem ser preenchidos de modo pertinente. Nesse caso, a veiculação de dados por meio de vocabulários de dados externos pode auxiliar no enriquecimento de dados.

O uso de contexto semântico por meio de vocabulários abertos para enriquecer dados existentes é uma prática comum na *Web Semântica* e já realizada pela *Europeana*. Para tanto, o uso de vocabulários abertos associados a soluções automatizadas é considerada uma maneira de melhorar a qualidade dos dados. Contudo, a grande questão é apoiar e incentivar às instituições provedoras para que contribuam com essa demanda. Além disso, há a questão da divergência entre os vocabulários, pois existem vários vocabulários criados e mantidos pelos provedores e outros pelos agregadores, que em geral não são publicados de maneira a serem reutilizados (EUROPEANA PRO, 2015).

Ainda a respeito dessa questão, existe o problema da relutância em reutilizar ou vincular dados que não são conhecidos. Sendo assim, a questão da proveniência dos dados é um fator determinante quando se propõe a vinculação e o enriquecimento de dados com conjuntos de dados externos. Arakaki (2019) destaca a importância da proveniência não só na questão da publicação dos dados, mas também para o reuso de dados, levando em conta a confiança e a compreensão das informações.

O uso de modelos para fornecer informações sobre proveniência e manutenção de dados, como a *W3C Provenance ontology* (PROV-O) pode auxiliar nessa questão (EUROPEANA PRO, 2015). Segundo o WORLD WIDE WEB CONSORTIUM (2013b, não paginado, tradução nossa) a ontologia PROV (PROV-O) expressa o modelo de dados PROV, que por sua vez:

[...] fornece um conjunto de classes, propriedades e restrições que podem ser usadas para representar e trocar informações de proveniência geradas em diferentes sistemas e em diferentes contextos. Também pode ser especializado para criar novas classes e propriedades para modelar informações de proveniência para diferentes aplicativos e domínios.

Além do uso de ferramentas como os modelos para auxiliar na tarefa de tornar os dados mais confiáveis, o próprio W3C desenvolveu em 2017 um conjunto de melhores práticas, denominado *Data on the Web Best Practices*, com o intuito de tornar os recursos

disponíveis na *Web* mais confiáveis e preparados para serem reutilizados. O documento apresenta diversas práticas que são recomendadas para ajudar editores e consumidores de dados a lidar com os diferentes desafios enfrentados ao publicar e consumir dados na *Web* (WORLD WIDE WEB CONSORTIUM, 2017).

Dentro dessa perspectiva, os participantes da conferência enfatizaram que a qualidade dos dados pode ser afetada por sua publicação. Sendo assim, entende-se que os dados perdem qualidade quando não são adequados ou compatíveis com os propósitos a que foram destinados.

Por isso, a publicação de dados deve ser vista como parte do processo de qualidade dos dados (EUROPEANA PRO, 2015), o que implica em um gerenciamento que considere que os dados sejam aprimorados antes de sua publicação, ou seja, no momento do seu planejamento (Conceptualização). Neste planejamento, considerar-se o trabalho sinérgico entre modelos conceituais, ontologias e metadados.

Além disso, é importante que esse gerenciamento compreenda também os dados já publicados, após serem reutilizados e aprimorados por terceiros, pois estes podem ser reaproveitados para enriquecer dados antigos ou até mesmo preencher lacunas de dados ausentes.

Uma terceira questão discutida se refere à participação da comunidade usuária para medir a qualidade dos dados. Na curadoria digital é fundamental que o gerenciamento dos dados seja feito tendo em vista a comunidade a que são destinados. O que leva a mais uma das questões levantada na conferência, que foi a respeito dos objetivos a que os dados são endereçados e o fato de que tais objetivos mudam em diferentes contextos.

Na conferência, os participantes discutiram que os dados no patrimônio cultural são criados principalmente para fins internos e descritivos, o que os tornam bastante claros e pertinentes para os profissionais que criaram os dados. Porém, para os usuários finais, esses dados podem ser menos compreensíveis. Portanto, o problema nesse caso é de que os objetivos dos dados não são documentados o suficiente, dificultando a avaliação do grau de qualidade necessário para atender aos requisitos em uma determinada aplicação (EUROPEANA PRO, 2015).

Considerando as questões levantada durante a conferência da *EuropeanaTech*, bem como o estudo realizado até então, destaca-se que garantir que os dados sejam autênticos, confiáveis e utilizáveis, mantendo sua integridade não é uma tarefa simples, principalmente em ambientes abertos e com estruturas variáveis, como é o caso da *Web*. Por essa razão, a curadoria digital faz uso de um conjunto de habilidades e competências voltadas ao

gerenciamento contínuo dos dados que se mostram grandes aliadas desses ambientes informacionais, como a *Europeana*, que busca as possibilidades do ambiente *Web* para ampliar suas formas de divulgação e multidimensionar o uso dos recursos informacionais.

Retomando a proposta do modelo *Extended Digital Curation Lifecycle Model* (DCC&U) de acrescentar à curadoria digital uma camada adicional de representação dos aspectos semânticos dos recursos informacionais, percebe-se que a ênfase está em considerar além dos metadados que garantem a existência dos recursos, também os que garantem a persistência e a validade de seu domínio de conhecimento, transmitindo o seu contexto semântico em qualquer ambiente informacional.

Isso não significa que anteriormente os aspectos semânticos não eram considerados. O próprio conceito de metadados já implica na ideia de agregar semântica aos dados. Além disso, como visto, elementos descritivos sempre estiveram ligados a atividade curatorial. A questão-chave, contudo, é que na medida em que aumentam os processos e demandas computacionais, é necessário fornecer continuamente um maior nível semântico aos dados. É o que a *Europeana* tem tentado fazer em sua estrutura de organização, utilizando para isso os princípios e as tecnologias da *Web Semântica*, bem como os instrumentos de organização e representação da informação e do conhecimento. Apesar disso, há ainda questionamentos sobre a eficácia do seu modelo de dados, o EDM, principalmente devido à complexidade e ao alto custo que envolvem sua atividade.

A diversidade, a heterogeneidade e a qualidade variável dos metadados que são inseridos ao EDM gera diversos desafios para garantir uma representação satisfatória que assegure a contextualização semântica dos recursos informacionais do patrimônio cultural ao mesmo tempo em que mantenha a integridade e a atualização dos dados à medida que as fontes mudam ao longo do tempo. Como já discutido, a *Europeana* continuamente dispense esforços e investimentos para que esse setor entregue resultados satisfatórios a sua comunidade de usuários.

Por essa razão, como já discutido por Harvey (2010), entende-se que os metadados são elementos que contribuem para que a curadoria digital seja complexa e de custo alto. Configurando-se como estruturas complexas, os metadados são muitas vezes caros para se criar e manter por exigirem alto nível de especialização e ferramentas adequadas (GILLILAND, 2016). Portanto, entende-se que a curadoria digital é afetada em sua usabilidade e em seus modelos de custos. Como visto na seção três, cada uma de suas ações implica em um conjunto de diretrizes que envolve diferentes atores e ferramentas. Dentro dessas diretrizes devem ser considerados também os custos para a criação, o armazenamento e

a preservação, bem como para manter o acesso e assegurar as questões legais e de direitos autorais (HARVEY, 2010, CORUJO; DA SILVA; REVEZ, 2016).

Essas questões potencializam a dificuldade do emprego da prática da curadoria digital exigindo um significativo investimento em tecnologia e experiência profissional. Além disso, sendo a curadoria digital uma atividade de longo prazo, é necessário ter os requisitos adequados para garantir um processo contínuo, que mantenha-se constantemente atualizado em relação às melhores práticas e às tecnologias vigentes.

Para algumas instituições, esse pontos acabam por tornar a implementação de um programa de curadoria digital inviável, especialmente porque suas maiores vantagens são percebidas a longo prazo (ABBOTT, 2008). Em relação a *Europeana*, a amplitude de seu empreendimento exige que os dados que são fornecidos pelos provedores de conteúdo atendam a critérios específicos para garantir a qualidade e a sustentabilidade dos dados a longo prazo.

Por conta disso, muitas instituições que não conseguem alcançar esses critérios, não têm seus dados incorporados à plataforma, não tendo a oportunidade de ampliar a divulgação e disseminação de seus acervos e coleções. Esse é uma realidade que a *Europeana* busca constantemente mudar em suas estratégias, procurando formas mais baratas de tratar e manter o conteúdo digital (EUROPEANA STRATEGY 2015-2020, 2017).

Apesar dessas dificuldades, ficam evidentes os benefícios que a curadoria digital proporciona. Entre esses benefícios destacam-se:

- a) Acesso persistente a dados confiáveis com maior qualidade;
- b) Possibilidade de constante verificação da autenticidade, melhorando a confiabilidade dos dados;
- c) Uso de padrões e melhores práticas que permitem a interoperabilidade técnica, funcional dos aspectos semânticos dos dados;
- d) Garantia dos aspectos semânticos do contexto e da proveniência dos dados;
- e) Garantia do valor a longo prazo dos dados como potencial de evidência, garantindo que estejam disponíveis para reutilização;
- f) Ampliação das possibilidade de compartilhamento e acesso, multidimensional as formas de uso, reuso e análise dos dados ao longo do tempo;
- g) Preservação em relação a obsolescência tecnológica e a mudanças institucionais garantindo que os dados permaneçam significativos para a comunidade de usuário.

Como visto, todos esses benéficos estão condicionados a quantidade e a qualidade dos metadados que são fornecidos para cada processo. Portanto, retoma-se que os metadados são fatores essenciais para empreender a curadoria digital, não só porque eles permitem explicitar as definições semânticas necessárias ao seu funcionamento, mas principalmente porque são esses elementos que vão determinar as condições e os resultados da curadoria digital.

A próxima seção apresenta as considerações finais, pontuando os resultados obtidos, as reflexões alcançadas com este estudo, bem como os possíveis direcionamentos futuros.

5 CONSIDERAÇÕES FINAIS

A curadoria digital surge e se estabelece diante das possibilidades de organizar, de compartilhar e de processar os recursos informacionais em meio digital e, concomitantemente, como um campo de estudo e prática interdisciplinar que visa, além da preservação e do acesso contínuo em ambientes digitais, também a garantia de autenticidade, de confiabilidade, de integridade e de usabilidade dos dados. Para garantir tais objetivos, a curadoria digital apoia-se na representação proveniente dos metadados, a fim de compreender e de processar o conteúdo digital, tornando os metadados elementos essenciais em seu ciclo de ações, como discutido por Higgins (2007; 2008; 2011), Harvey (2010), Glushko (2013), Chao (2014), Oliver e Harvey (2016), entre outros.

Nesse sentido, a literatura ressalta que a necessidade dos metadados para o gerenciamento do conteúdo digital vai além da representação para identificação e recuperação, envolvendo principalmente a garantia de persistência nos ambientes digitais; o que, por sua vez, envolve sustentabilidade e interoperabilidade de dados. Tais conceitos ganharam importância significativa diante do volume e da complexidade dos dados. No atual cenário, que é marcado pelo uso intensivo das tecnologias emergentes, a capacidade dos dados de persistir com qualidade a longo prazo e ao mesmo tempo interoperar com diversas aplicações, passa a ser determinante para o sucesso de um ambiente/sistema informacional digital.

A análise da literatura aponta para mais uma função na qual os metadados são essenciais, quando o objetivo é assegurar além do uso e da preservação momentâneos. Essa função é a de agregar valor aos dados, de modo a conferir contexto semântico e garantir a adequação ao propósito informacional em diversos contextos de uso (RUSBRIDGE *et al.*, 2005; DALLAS, 2007; CONSTANTOPOULOS; DALLAS, 2008; CONSTANTOPOULOS; DALLAS *et al.*, 2009).

Seguindo essa perspectiva, a pesquisa teve como objetivo geral apresentar um panorama das ações que caracterizam a curadoria digital de modo a identificar como os metadados influenciam sua prática, considerando a representação de dados no ambiente *Web*.

Para tanto, as seções dois e três apresentam uma revisão de literatura que contempla os dois objetos de pesquisa: os metadados e a curadoria digital. As discussões foram feitas levando em consideração a necessidade de refletir no âmbito da Ciência da Informação os procedimentos e os métodos envolvidos no tratamento informacional, em especial, no

tratamento dos dados que estão alocados em ambientes informacionais direcionados pela construção semântica e pela ligação entre os dados, como no ambiente *Web*.

Como resultado, foi possível contemplar o primeiro objetivo específico definido para pesquisa: descrever os fundamentos dos metadados, da representação da informação no ambiente *Web* e da curadoria digital. Para isso, os conceitos de dados e de metadados foram discutidos (SOUZA; ALVARENGA, 2004, SANTOS; SANT'ANA, 2013; POMERANTZ, 2015), bem como os aspectos da representação da informação no ambiente *Web* (SANTOS; ALVES, 2009, POMERANTZ, 2015, GILLILAND, 2016, ZENG; QIN, 2016), resultando em uma breve análise das tecnologias semânticas.

Os pontos abordados na segunda seção tiveram o intuito de elucidar como está o cenário das pesquisas que envolvem os metadados, considerando as novas dimensões do tratamento informacional. Observou-se que, mesmo com uma série de tecnologias trabalhando de modo integrado para que os recursos digitais sejam bem representados, o objetivo de alcançar um tratamento informacional eficaz ainda se mantém. A grande questão, no entanto, é que esse tratamento precisa levar em consideração estruturas padronizadas, recuperáveis e processáveis principalmente de modo automatizado, o que, como fica claro na literatura, exige um alto nível de descrição proveniente dos diversos tipos de metadados.

Em relação à curadoria digital, observou-se na terceira seção que o número de estudos na temática não é muito expressivo, principalmente no Brasil. No entanto, identificaram-se alguns núcleos de pesquisa que são bastante atuantes na área, como o *Digital Curation Centre* (DCC), no Reino Unido e o *Digital Curation Unit* (DCU), Grécia. No Brasil, os autores que mais trabalham com a temática são Sayão (2012; 2016); Sayão e Sales (2012; 2016); Siebra *et al.* (2013); Siebra, Borba e Miranda (2016); Santos (2014); Silva (2017); Jorente, Silva e Pimenta (2015), e Marcondes (2016; 2018).

Observou-se que os primeiros trabalhos que utilizaram o termo 'curadoria digital' foram apresentados no início do século XXI, havendo um crescimento médio constante nos últimos anos. No entanto, nem sempre o termo 'curadoria digital' é utilizado para descrever a prática de gerenciamento e de manutenção em meio digital. Termos como 'curadoria de dados', 'curadoria de dados de pesquisa', 'curadoria e preservação', 'curadoria de conteúdo digital' aparecem com frequência. Além disso, há uma parcela desses trabalhos que estuda a prática da curadoria digital em contextos específicos, como em repositórios institucionais ou em coleções e acervos temáticos e arquivísticos.

Os usos desses termos em português permitem traçar um esboço sobre os rumos que a pesquisa em curadoria digital tem tomado no Brasil, estando fortemente voltada para a gestão

de dados, no contexto das pesquisas acadêmicas e científicas, ou ainda, envolvendo casos específicos de tratamento de acervos e coleções. Tal constatação resultou em uma das justificativas dessa pesquisa, uma vez que são menos expressivos os estudos que envolvem a curadoria digital no universo dos patrimônios culturais. Em contrapartida, no cenário internacional, há um maior número de projetos e de iniciativas que consideram os aspectos da curadoria digital no âmbito da gestão dos patrimônios culturais, envolvendo principalmente as questões de integração de acervos heterogêneos no ambiente *Web*.

Devido à variedade de temas a que a curadoria digital costuma ser vinculada, não é simples definir quais são seus objetivos e as ações que caracterizam sua prática. No entanto, a análise da literatura revela que ao longo dos anos diversos autores têm tentado transmitir uma conceituação que defina os reais propósitos da curadoria digital, o que é apresentado em forma de quadro na terceira seção (Quadro 6). Por meio da análise comparativa entre as definições apresentadas pelos autores selecionados foi possível ressaltar algumas interpretações que auxiliaram no objetivo proposto.

Destaca-se que a curadoria digital tem suas atividades orientadas às questões que afetam o ciclo de vida de um recurso informacional, envolvendo desde a criação, a aquisição, o processamento, a manutenção, até o acesso e a disponibilização dos dados.

A curadoria digital considera como objeto central de suas ações os dados que, para a maioria dos autores, inclui tudo aquilo que é armazenado como fluxos de *bits*, compondo os denominados objetos digitais ou ativos digitais, que podem ser mais simples, como a expressão mínima de um conteúdo, ou mais complexos, envolvendo grandes conjuntos de dados.

Independente da complexidade dos dados, concluiu-se que o foco da curadoria digital está no valor informacional que esses dados possuem e que, portanto, suas ações apresentam direta relação com a manutenção da concepção intelectual, envolvendo, assim, além das ações de arquivamento e de preservação digital, também as ações para manter e agregar valor por meio da representação da informação. Desse modo, constatou-se que o objetivo da curadoria digital é garantir o valor dos recursos informacionais para uso atual e futuro, para que eles possam não só se manter acessíveis, mas com qualidade e confiabilidade para serem reaproveitados e reutilizados em diversos contextos.

Infere-se, então, que o objetivo da curadoria digital não é algo simples, mas sim que se ramifica em diversos objetivos específicos para atingir um objetivo total. Por isso, sua abordagem em geral é apresentada de forma cíclica, em formato de modelos baseados em

estratégias, métodos e tecnologias, que contemplam todos os estágios de vida e os relacionamentos dos dados.

Tendo isso em conta, o segundo objetivo específico contemplado foi analisar os modelos de curadoria digital, buscando elucidar quais seriam as ações e os objetivos associados a sua prática. Os modelos foram escolhidos considerando aqueles que não eram restritivos a coleções e sistemas específicos, de modo a verificar seus aspectos de forma geral e abrangente.

Como resultado, foram apresentados quatro modelos que são considerados pela literatura como referência na temática e possuem relevância para o propósito dessa pesquisa. São eles: o modelo de referência *Open Archival Information Systems* (OAIS), o modelo do *Joint Information Systems Committee* (JISC), o modelo do *Digital Curation Centre* (DCC) e o modelo o *Extended Digital Curation Lifecycle Model* do DCU (DCC&U). A análise comparativa desses modelos demonstrou que existem ações que são recorrentes, mesmo quando apresentam nomenclaturas diferentes. Tais ações podem ser consideradas como um denominador comum entre os modelos apresentados, o que leva a refletir que existe um conjunto de ações mínimas que devem ser desempenhadas para que a curadoria digital seja efetiva. Essas ações envolvem principalmente o objetivo de manter a informação com qualidade e caráter inalterado ao longo do tempo (GLUSHKO, 2013), o que reforça a ideia de que a curadoria digital está diretamente relacionada com a representação da informação.

As discussões oferecidas pelas seções dois e três deram base para responder a questão proposta. Considerando que os metadados são intrínsecos ao processo de curadoria independente do seu âmbito de aplicação, questionavam-se como os metadados influenciam as ações da curadoria digital.

A resposta foi apresentada na quarta seção, na qual discutiu-se a relação entre os metadados e a curadoria digital, tendo como base os princípios da Organização e Representação da Informação e do Conhecimento advindos da Biblioteconomia e Ciência da Informação.

Verificou-se que, assim como a Biblioteconomia e a Ciência da Informação fazem uso dos princípios da Organização e Representação para o tratamento da informação, na curadoria digital, os mesmos princípios norteiam suas ações. Cada uma dessas ações requer um conjunto de dados sobre os dados que estão sendo gerenciados. Esses dados são os metadados, que em geral assumem uma estrutura padronizada, compondo um padrão ou esquema de metadados. Para tanto, existem alguns padrões que costumam ser usados para auxiliar as ações da curadoria digital, como os padrões de preservação e de proveniência, bem

como os padrões de domínios específicos, buscando fornecer metadados adequados para cada função.

Sendo assim, constatou-se que os metadados, bem como os padrões de metadados, influenciam as ações da curadoria digital na medida em que esses elementos permitem explicitar as definições semânticas que são necessárias à comunicação dentro do processo de gerenciamento, principalmente de modo automatizado.

Ao passo que aumenta o número de processos automatizados, envolvendo a produção, a análise e o uso de dados e de informações, aumenta também a necessidade de instrumentos que assegurem a correta interpretação desses recursos digitais, tanto para os humanos, quanto para as máquinas. Portanto, ressalta-se a importância de que os processos de curadoria digital sejam aderentes a padrões, seguindo recomendações e princípios.

Além disso, observou-se que o fato da curadoria digital estar apoiada nos metadados implica também que ela é sustentada pelas novas dimensões que envolvem o tratamento da informação em ambientes digitais. Isso significa que a curadoria digital sofre influência das soluções tecnológicas que vem sendo apresentadas para o atual cenário.

Na quarta seção foi possível, também, trabalhar o terceiro objetivo específico proposto que era verificar como os metadados possibilitam assegurar o contexto semântico dos recursos informacionais do patrimônio cultural. Verificou-se que, os metadados possibilitam assegurar o contexto semântico dos recursos informacionais a partir das soluções tecnológicas que explicitam os significados e as relações dos dados relativos ao corpo de conhecimento a que esses recursos pertencem. Ou seja, a partir de ferramentas que permitam identificar os conceitos e as relações existentes entre os vários recursos informacionais disponibilizados, dando condições para que esses conceitos e relações sejam codificados e processados por computador.

Dentre as soluções estão às tecnologias semânticas criadas para permitir que a *Web* atinja um maior nível de interação semântica entre suas funções (BERNERS-LEE; HENDLER; LASSILA, 2001). Tais tecnologias, como *Extensible Markup Language* (XML), *Resource Description Framework* (RDF), *Simple Knowledge Organization System* (SKOS), *Uniform Resource Identifier* (URI), *Ontology Web Language* (OWL) e suas derivações, permitem a organização e a representação da informação e do conhecimento para as máquinas, a partir de padrões de metadados e ontologias.

Assim como os padrões de metadados que são criados para especificar as definições semânticas dos seus elementos descritivos em formas padronizadas (HIGGINS, 2007a), as ontologias são desenvolvidas para explicitar um modelo de dados que representa um conjunto

de conceitos e relacionamentos dentro de um domínio (RAMALHO, 2006; FUSCO, 2010). Isso possibilita compartilhar uma determinada perspectiva acerca de um campo de conhecimento por meio de um vocabulário comum (SALES; CAFÉ, 2009) e ainda integrar estruturas de metadados para obter interoperabilidade entre sistemas heterogêneos, o que é significativo para a curadoria digital, pois permite adequar os dados e as informações a cada finalidade e contexto de uso. Ao representarem formalmente os elementos de um corpo do conhecimento, as ontologias permitem assegurar que o contexto semântico a que um recurso informacional digital pertence seja explicitado e mantido durante e o seu ciclo de vida.

Atuando em conjunto às ontologias para proporcionar um maior nível semântico, encontram-se os princípios do *Linked Open Data* (LOD). Para a curadoria digital, principalmente de recursos do patrimônio cultural, a possibilidade de reutilizar dados já publicados e vinculados otimiza a representação informacional e, ao mesmo tempo, amplia as possibilidades de conectar e atribuir metadados apropriados a partir dos relacionamentos semânticos entre os dados, que para Marcondes (2018) são os relacionamentos culturalmente significativos.

Diante disso, buscou-se compreender como essa interação entre as tecnologias semânticas, os metadados e a curadoria digital ocorre na prática, cumprindo, assim o último objetivo específico estabelecido. Considerando o posicionamento de Sayão (2016) que defende que a curadoria digital viabiliza uma gestão dinâmica de coleções digitais que é favorável ao desenvolvimento de serviços *online* e espaços de interação em torno dos acervos digitais heterogêneos, apresentou-se um caso que se enquadra nas tendências de organização e de representação em ambientes digitais, que levam em consideração os aspectos semânticos e a ligação entre os dados.

Esse caso é o da comunidade *Europeana*, que mantém uma plataforma destinada a reunir e a estruturar dados de diversas instituições europeias que trabalham com patrimônio cultural, almejando disponibilizar seus acervos e coleções de forma aberta, promovendo amplo reuso e o acesso dos dados. Para atingir esse objetivo, a plataforma executa um detalhado programa de contextualização semântica que tem como base o modelo de dados denominado *Europeana Data Model* (EDM), que é fundamentado nas tecnologias semânticas e em padrões e em princípios, permitindo agregar e reutilizar dados.

Por meio do EDM, a *Europeana* coordena todo o fluxo de dados que interage na plataforma, permitindo o seu gerenciamento. Esse gerenciamento se apresenta em uma estrutura de três níveis, que considera respectivamente os dados, os instrumentos para padronizar e enriquece-los, permitindo o acesso e as necessidades das comunidades que fazem

uso da plataforma. Com esses três níveis, a *Europeana* demonstra sua grande preocupação não só em garantir o planejamento e a execução de suas ações, mas principalmente em garantir que os dados que interagem na plataforma mantendo sua qualidade e confiabilidade para o acesso, uso e reuso a longo prazo. Portanto, percebe-se que existem grandes correspondências no modo como a plataforma *Europeana* executa seu gerenciamento de dados e as ações consideradas necessárias pela curadoria digital, não só em relação à questão da manutenção e à preservação dos dados no interior de seu sistema, mas principalmente em relação ao processo de agregar valor, tarefa que é desempenhada pelo EDM.

As discussões apresentadas no estudo permitem uma reflexão geral, demonstrando que a curadoria digital, apesar de se diferenciar em diversos aspectos da tradicional curadoria desempenhada em instituições de patrimônios culturais, ainda mantém os fundamentos de sua origem. Esses fundamentos estão centrados na tarefa de garantir o propósito informacional dos recursos por meio da contextualização semântica. Essa contextualização semântica, que tem por base o corpo de conhecimento de um recurso informacional, destina-se a orientar as ações que serão realizadas sobre o acervo ou coleção, sejam eles analógicos ou digitais.

Tais ações, por sua vez, são tradicionalmente realizadas por um o curador, que é um especialista de domínio que detém o conhecimento de um determinado conteúdo. Na curadoria digital, no entanto, a função do curador não é necessariamente desempenhada por um especialista de domínio, mas sim por um conjunto de tecnologias, princípios e profissionais que capacitam máquinas para exercerem em parte essa função.

Para o estabelecimento da curadoria digital seja possível, é necessário que o conhecimento de domínio seja organizado e representado também para as máquinas, o que implica no uso de ferramentas que explicitem a dimensão semântica dos dados para processamento de modo automatizado. Como visto, essa não é um tarefa simples, principalmente porque implica em adicionar uma série camadas de metadados com diferentes tipologias e funções para seja possível garantir a autenticidade, a confiabilidade, a integridade e a usabilidade de modo a ter dados interoperáveis com qualidade e consistência. Isso acaba por potencializar os desafios do gerenciamento de dados, contribuindo para que a curadoria digital seja uma atividade complexa e de custo alto.

Apesar disso, vale ressaltar os benefícios que a curadoria digital proporciona, garantindo acesso persistente, verificação de autenticidade, diretrizes e melhores práticas, contextualização semântica, proveniência dos dados e o potencial como evidência. Além disso, ainda é possível citar a possibilidade de multidimensionar as formas de uso, reuso e

análise dos dados ao longo do tempo em oposição à obsolescência tecnológica e a mudanças institucionais.

Diante do exposto e dos resultados alcançados, observa-se que muitos estudos ainda são necessários para o satisfatório aprofundamento no universo que envolve a curadoria digital, em especial, no que diz respeito às tecnologias que apoiam a ampla gama de aspectos que devem ser considerados para um efetivo gerenciamento de dados. Ainda, é possível considerar a constante atualização das tecnologias e a interdisciplinaridade que envolve esse universo, o que incita sempre novos estudos.

Sendo assim, é possível apontar alguns direcionamentos futuros:

- Aprofundamento dos estudos que relacionam a curadoria digital ao *Linked Open Data*;
- Ampliação do escopo da pesquisa para demais iniciativas como a *Digital Public Library of America* (DPLA) e a *DigitalNZ*;
- Aprofundamento no estudo e na aplicação do *Extended Digital Curation Lifecycle Model* (DCC&U) em diferentes ambientes informacionais;
- Desenvolvimento de um estudo de caso e de uma pesquisa aplicada na *Europeana*, oportunizando aprofundar os conhecimentos sobre seu sistema.

Como foi possível perceber com o desenvolvimento dessa dissertação, mesmo com a complexidade dos processos que envolvem as tecnologias semânticas e a própria abrangência conceitual da curadoria digital, os assuntos abordados não devem deixar de ser estudados e discutidos no âmbito da Ciência da Informação.

REFERÊNCIAS

- ABBOTT, D. **What is digital curation?** Edimburgo: Digital Curation Centre, 2008. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation>. Acesso em: 30 jun. 2019.
- ALMEIDA, M. B. Uma introdução ao XML, sua utilização na internet e alguns conceitos complementares. **Ci. Inf.**, Brasília, v. 31, n. 2, p. 5-13, 2002. Disponível em: <http://www.scielo.br/pdf/ci/v31n2/12903>. Acesso em: 30 jun. 2019.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 134f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rachel.pdf. Acesso em: 30 jun. 2019.
- ALVES, R. C. V. Metadados para representação e recuperação da informação em ambiente web. **Seminário Serviços de Informação em Museus**, São Paulo, SP, p. 95-106, 2016. Disponível em: <http://biblioteca.pinacoteca.org.br:9090/publicacoes/index.php/sim/article/view/118/123>. Acesso em: 30 jun. 2019.
- ANTONIOU, G. *et al.* **A semantic web primer**. 3.ed. Cambridge, Massachusetts; London, England : The MIT Press, 2004. Disponível em: [http://prof.mau.ac.ir/images/Uploaded_files/A%20Semantic%20Web%20Primer-The%20MIT%20Press%20\(2012\)\[7460174\].PDF](http://prof.mau.ac.ir/images/Uploaded_files/A%20Semantic%20Web%20Primer-The%20MIT%20Press%20(2012)[7460174].PDF). Acesso em: 30 jun. 2019.
- ARAKAKI, F. A. **Linked data**: ligação de dados bibliográficos. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016. Disponível em: <http://hdl.handle.net/11449/147979>. Acesso em: 10 jan. 2020.
- ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados**: modelo baseado na família PROV. 2019. 140 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016. Disponível em: <http://hdl.handle.net/11449/180490>. Acesso em: 10 jan. 2020.
- ARAKAKI, F. A.; SIMIONATO, A. C.; SANTOS, P. L. V. A. C. Integrando catálogos entre bibliotecas, arquivos, museus e galerias de arte: perspectiva da Europeia e da DPLA. **RBBB**. Revista Brasileira de Biblioteconomia e Documentação, v. 13, p. 2250-2268, 2017. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1016/937>. Acesso em: 10 jan. 2020.
- ARAÚJO, C. A. A. Fundamentos da ciência da informação: correntes teóricas e o conceito de informação. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 4, n. 1, p. 57–79, 2014. Disponível em: <http://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/19120/10827>. Acesso em: 30 jun. 2019.
- ARAÚJO, C. A. A. Teorias e tendências contemporâneas da Ciência da Informação. **Informação em Pauta**, Fortaleza, v. 2, n. 2, p. 9-34, 2017. Disponível em:

<http://www.periodicos.ufc.br/informacaoempauta/article/view/20162>. Acesso em: 30 jun. 2019.

ASSUMPTÃO, F. S.; SANTOS, P. L. V. A. C. Representação no domínio bibliográfico: um olhar sobre os Formatos MARC 21. **Perspectivas em Ciência da Informação**, v.20, n.1, p.54-74, jan./mar. 2015. Disponível em: <https://doi.org/10.1590/1981-5344/2054>. Acesso em: 30 jun. 2019.

BAKER, T. *et al.* **Library linked data incubator group Final Report**. W3C Incubator Group Report, 2011. Disponível em: <http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/>. Acesso em: 30 jun. 2019.

BAKER, T. Libraries, languages of description, and linked data: A Dublin Core perspective. **Library Hi Tech**, v. 30, n. 1, Mar, 2012. p. 116-133. Disponível em: <http://www.emeraldinsight.com/doi/abs/10.1108/07378831211213256/>. Acesso em: 20 ago. 2015. DOI: 10.1108/07378831211213256

BALL, A. **Review of the state of the art of the digital curation of research data**. University of Bath: Bath, 2010. Disponível em: [erim1rep091103ab11.pdf](http://www.bath.ac.uk/ims/rim1rep091103ab11.pdf) . Acesso em: 30 jun. 2019.

BALL, A. **Review of data management lifecycle models**. University of Bath, IDMRC: Bath 2012. Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/206543/redm1rep120110ab10.pdf>. Acesso em: 30 jun. 2019.

BAPTISTA, D. M. Impacto dos metadados na representação descritiva. **Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis**, v.12, n.180 2, p. 177-190, jul./dez., 2007. Disponível em: <https://revista.acbsc.org.br/racb/article/view/529/663>. Acesso em: 30 jun. 2019.

BEAGRIE, N. The continuing access and digital preservation strategy for the UK Joint Information Systems Committee (JISC). **D-Lib**, v.10, n.7/8. 2004a. Disponível em: <http://www.dlib.org/dlib/july04/beagrie/07beagrie.html>. Acesso em: 30 jun. 2019.

BEAGRIE, N. The Digital Curation Centre. **Learned publishing**, v. 17, n. 1, p. 7-9, 2004b. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1087/095315104322710197>. Acesso em: 30 jun. 2019.

BEAGRIE, N. Digital curation for science, digital libraries, and individuals. **The International Journal of Digital Curation**, Edinburgh, v. 1, Iss. 1, 2006. Disponível em: <http://www.ijdc.net/article/view/6>. Acesso em: 30 jun. 2019.

BEAGRIE, N., POTHEEN, P. The digital curation: digital archives, libraries and e-science seminar. **Ariadne**, Iss., 30, 2001. Disponível em: <http://www.ariadne.ac.uk/issue/30/digital-curation/>. Acesso em: 30 jun. 2019.

BERNERS-LEE, T. **Weaving the web: the past, present and future of the World Wide Web by its Inventor**. London: Texere, 2000.

BERNERS-LEE, T. **Semantic web road map**. World Wide Web Consortium (W3C). 2001. Disponível em: <http://www.w3.org/DesignIssues/Semantic.html>. Acesso em: 30 jun. 2019.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, 2001, p. 29-37. Disponível em: <https://www.ida.liu.se/~evabl45/files/bernersLee01Semantic.pdf>. Acesso em: 30 jun. 2019.

BISHOP, B. W.; HANK, C. Measuring FAIR Principles to inform fitness for use. **The International Journal of Digital Curation**, Edinburgh, v.13, Iss. 1, 2018. Disponível em: <http://www.ijdc.net/article/view/630/514>. Acesso em: 30 jun. 2019.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data: the story so far. *In*: SHETH, A. P. (Ed.) **Semantic services, interoperability and web applications: emerging concepts**. Hershey: Information Science Reference, 2009. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 30 jun. 2019.

CAMPOS, J. A. G. **Análise conceitual sobre as relações semânticas em Ciência da Informação**: contribuições para o desenvolvimento de ontologias. 2009. 135f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: <http://hdl.handle.net/1843/ECID-7V4PLA>. Acesso em: 10 jan. 2020.

CASTRO, F. F. de. **Elementos de interoperabilidade na catalogação descritiva**: configurações contemporâneas para a modelagem de ambientes informacionais digitais. 201f. Tese (Doutorado em Ciência da Informação) -- Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012. Disponível em: http://www.marilia.unesp.br/Home/PosGraduacao/CienciadaInformacao/Dissertacoes/Castro,%20F.F._doutorado_CI_2012.pdf. Acesso em: 10 jan. 2020.

CASTRO, F. F.; SIMIONATO, A. C.; ZAFALON, Z. R. Aspectos relacionais entre ontologia e metadados: considerações interdisciplinares. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17, Salvador, 2016. **Anais[...]**. Salvador: Universidade Federal da Bahia, 2016. Disponível em: http://www.academia.edu/download/51348047/2016_-_Enancib_-_Aspectos_relacionais_entre_ontologias_e_metadados.pdf. Acesso em: 30 jun. 2019.

CHAO, T. C. Enhancing metadata for research methods in data curation. *In*: American Society for Information Science and Technology - ASIST, 77. Seattle, WA, 2014. **Proceedings [...]**. Seattle, WA: [s.n], v.51, n.1 p.1-4. 2014. Disponível em: <https://doi.org/10.1002/meet.2014.14505101103>. Acesso em: 10 jan. 2020.

CHARLES, V.; ISAAC, A. Enhancing the Europeana Data Model (EDM). **Europeana Foundation**, 2015. Disponível em: https://pro.Europeana.eu/files/Europeana_Professional/Publications/EDM_WhitePaper_17062015.pdf. Acesso em: 10 jan. 2020.

CLATIN, M. *et al.* Digital curators at work: analyzing emerging professional identities at the Bibliothèque nationale de France (BnF). *In*: IFLA WORLD LIBRARY AND

INFORMATION CONGRESS, 2014. **Proceedings** [...]. Lyon, FR: IFLA, 2014. Disponível em: <https://hal-bnf.archives-ouvertes.fr/hal-01098526/>. Acesso em: 30 jun. 2019.

CONCORDIA, C.; GRADMANN, S.; SIEBINGA, S. Not (just) a Repository, nor (just) a Digital Library, nor (just) a Portal: A Portrait of Europeana as an API. *In: WORLD LIBRARY AND INFORMATION CONGRESS IFLA GENERAL CONFERENCE AND COUNCIL*, 75, 2009. Milão, Itália. **Proceedings** [...] Milão, Itália: IFLA, 2009. Disponível em: <https://www.ifla.org/past-wlic/2009/193-concordia-en.pdf>. Acesso em: 30 jun. 2019.

CONEGLIAN, C. S. *et al.* Semantic web technologies in information architecture. **Revista Interamericana de Bibliotecología**, v. 42, n. 1, p. 23-35, 2019. Disponível em: <http://dx.doi.org/10.17533/udea.rib.v42n1a03>. Acesso em: 30 jun. 2019.

CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E. Europeana no linked open data: conceitos de web semântica na dimensão aplicada das Humanidades Digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 22, n. 48, p. 88-99, 2017. Disponível em: <https://www.redalyc.org/pdf/147/14748878008.pdf>. Acesso em: 30 jun. 2019.

CONSTANTOPOULOS, P.; DALLAS, C. Aspects of a digital curation agenda for cultural heritage. *In: IEEE International Conference on Distributed Human-Machine Systems*, 2008, p. 1-6. **Proceedings** [...]. Athens, Greece: IEEE, 2008. Disponível em: <http://www.dcu.gr/wp-content/uploads/2016/10/Aspects-of-a-digital-curation-agenda-for-cultural-heritage.pdf>. Acesso em: 30 jun. 2019.

CONSTANTOPOULOS, P. *et al.* DCC&U: An extended digital curation lifecycle model. **International Journal of Digital Curation**, Edinburgh, v. 4, n. 1, 2009. Disponível em: <https://doi.org/10.2218/ijdc.v4i1.76>. Acesso em: 30 jun. 2019.

CONWAY, P. **Preservação no Universo Digital**. Rio de Janeiro: Arquivo Nacional, 2001. Disponível em: <http://arqsp.org.br/wp-content/uploads/2017/07/52.pdf>. Acesso em: 30 jun. 2019.

CORUJO, L; DA SILVA, C. G.; REVEZ, J. Digital curation and costs: approaches and perceptions. *In: International Conference on Technological Ecosystems for Enhancing Multiculturality*, 4, 2016. **Proceedings** [...]. Salamanca, Spain: TEEM'16, 2016. p. 277-284. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3012430.3012529>. Acesso em: 08 jan. 2020.

CRUZ, I. F.; XIAO, H. The role of ontologies in data integration. **Engineering intelligent systems for electrical engineering and communications**, Bentley, WA, v. 13, n. 4, p. 245, 2005. Disponível em: <https://www.cs.uic.edu/~advis/publications/dataint/eis05j.pdf>. Acesso em: 30 jun. 2019.

DALLAS, C. An Agency-oriented Approach to Digital Curation Theory and Practice. *In: INTERNATIONAL CULTURAL HERITAGE INFORMATICS MEETING (ICHIM07)*, Toronto, 2007. **Proceedings** [...]. Toronto: Archives & Museum Informatics, 2007. Disponível em: <http://www.archimuse.com/ichim07/papers/dallas/dallas.html>. Acesso em: 30 jun. 2019.

DALLAS, C. Digital curation beyond the “wild frontier”: a pragmatic approach. **Archival Science**, v. 16, n. 4, p. 421-457, 2016. Disponível em: <https://doi.org/10.1007/s10502-015-9252-6>. Acesso em: 30 jun. 2019.

DATAONE. **Best Practices**, 2019. Disponível em: <https://www.dataone.org/best-practices>. Acesso em: 31 jun. 2019.

DIGITAL CURATION CENTRE. **History of the DCC**, 2004-2019a. Disponível em: <http://www.dcc.ac.uk/about-us/history-dcc/history-dcc>. Acesso em: 30 jun. 2019.

DIGITAL CURATION CENTRE. **DCC Curation Lifecycle Model**, 2004-2019b. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 30 jun. 2019.

DIGITAL CURATION CENTRE. **What is representation information?** 2004-2019c. Disponível em: <http://www.dcc.ac.uk/node/9558>. Acesso em: 30 jun. 2019.

DIGITAL CURATION CENTRE. **DCC Charter and Statement of Principles**. 2004-2019d. Disponível em: <http://www.dcc.ac.uk/about-us/dcc-charter/dcc-charter-and-statement-principles>. Acesso em: 30 jun. 2019.

DOERR, M. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. **AI magazine**, v. 24, n. 3, p. 75-75, 2003. Disponível em: <https://doi.org/10.1609/aimag.v24i3.1720>. Acesso em: 30 jun. 2019.

DOERR, M. **DCC Digital Curation Manual: Instalment on Ontologies**. Edimburgo, Escócia: Digital Curation Manual. 2008. Disponível em: <http://hdl.handle.net/1842/3341>. Acesso em: 10 jan. 2020.

DOERR, M. *et al.* The Europeana Data Model (EDM). *In: IFLA GENERAL CONFERENCE AND ASSEMBLY, 76, 2010, Gothenburg. Proceedings [...]* Gothenburg: IFLA, 2010. Disponível em: <https://www.ifla.org/past-wlic/2010/>. Acesso em: 10 jan. 2020.

DUBLIN CORE METADATA GLOSSARY. **Using Dublin Core**, 2001. Disponível em: <https://www.dublincore.org/specifications/dublin-core/usageguide/2001-04-12/glossary/>. Acesso em: 10 jan. 2020.

DUVAL, E. *et al.* Metadata Principles and Practicalities. **D-Lib**, v. 8, n. 4, 2002. Disponível em: <http://www.dlib.org/dlib/april02/weibel/04weibel.html>. Acesso em: 30 jun. 2019.

EUROPEANA. **Europeana Data Model Primer**. Dublin: Europeana Professional Website, 2013. Disponível em: https://pro.Europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf. Acesso em: 10 jan 2020.

EUROPEANA. **Definition of the Europeana Data Model v5.2.8**. 2017a. Disponível em: https://pro.Europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf. Acesso em: 10 jan 2020.

EUROPEANA. **Europeana Data Model – mapping guidelines v2.4**. 2017b. Disponível em: https://pro.Europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements

nts/EDM_Documentation/EDM_Mapping_Guidelines_v2.4_102017.pdf. Acesso em: 10 jan 2020.

EUROPEANA. **Europeana publishing guide v1.8**, 2019. Disponível em: <https://pro.europeana.eu/post/publication-policy>. Acesso em: 10 jan. 2020.

EUROPEANA COLLECTIONS. **Terms of use**, [2020]. Disponível em: <https://www.europeana.eu/portal/pt/rights/terms.html>. Acesso em: Acesso em: 10 jan 2020.

EUROPEANA PRO. **We want better data quality: NOW!**, 2015. Disponível em: <https://pro.europeana.eu/page/data-quality-etech15-roundtables>. Acesso em: 10 jan 2020.

EUROPEANA PRO. **Our mission: history**, 2017. Disponível em: <https://pro.europeana.eu/our-mission/history>. Acesso em: 10 jan 2020.

EUROPEANA PRO. **Europeana Rest API: Europeana Data Model**, [2020] <https://pro.europeana.eu/resources/apis/intro>. Acesso em: 10 jan 2020.

EUROPEANA STRATEGY 2015-2020. **Estratégia 2015-2020: 'we transform the world with culture'**, 2017. Disponível em: https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Strategy%202020.pdf. Acesso em: 10 jan 2020.

FAIR PRINCIPLES. **Fair Principles**, [2019]. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 30 jun. 2019.

FERREIRA, M. **Introdução à preservação digital: conceitos, idéias e actuais consensos**. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. 88 p. Disponível em: <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 30 jun. 2019.

FUSCO, E. **Modelos conceituais de dados como parte do processo da catalogação: perspectiva de uso dos FRBR no desenvolvimento de catálogos bibliográficos digitais**. 2010. 249f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2010. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/fusco_e_do_mar.pdf. Acesso em: 30 jun. 2019.

GARTNER, Richard. **Metadata**. New York, NY: Springer Berlin Heidelberg, 2016.

GIARETTA, D. **DCC Approach to Digital Curation**. Digital Curation Centre. Disponível em: <http://www.dcc.ac.uk/sites/default/files/documents/DCCApproachtoDigitalCuration-20040827.pdf>. Acesso em: 31 jun. 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4 ed. São Paulo: Atlas, 2008.

GILLILAND, A. J. Setting the stage. *In*: BACA, Murtha (Ed.). **Introduction to metadata**. Los Angeles, CA: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 30 jun. 2019.

GLUSHKO, R. J (Ed.). **The discipline of organizing**. Cambridge, Mass: The MIT Press, 2013, 540 p.

HARPRING, P. **Introdução aos vocabulários controlados**: terminologia para arte, arquitetura e outras obras culturais. São Paulo: ACAM Portinari, 2016. Disponível em: <https://www.sisemsp.org.br/wp-content/uploads/2013/12/Vocabularios%20Controlados%20-%20Digital.pdf>. Acesso em: 10 jun. 2020.

HARVEY, R. **Digital curation**: a how-to-do it manual. New York, NY: Neal-Schuman Publishers, Inc., 2010.

HASLHOFER, B.; ISAAC, A. data.europeana.eu: the europeana linked open data pilot. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2011, The Hague, Holland. **Proceeding [...]** The Hague, Holland: DC-at The Hague, 2011. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3625>. Acesso em: 10 jun. 2020.

HELLMAN, E. **Can librarians be put directly onto the semantic web?**. Go to Hellman blog, 2009. Disponível em: <https://go-to-hellman.blogspot.com/2009/08/can-librarians-be-put-directly-onto.html>. Acesso em: 10 jun. 2020.

HENNING, P. *et al.* Desmistificando os Princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *In*: XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XIX ENANCIB), 19, Londrina, PR, 2018. **Anais [...]**. Londrina, PR: Universidade Estadual de Londrina, 2018. Disponível em: <http://www.brapci.inf.br/index.php/res/v/103243>. Acesso em: 10 jun. 2020.

HIGGINS, S. **What are metadata standards?** Edimburgo: Digital Curation Centre, 2007a. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>. Acesso em: 30 jun. 2019.

HIGGINS, S. **Using metadata standards**. Edimburgo: Digital Curation Centre, 2007b. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/using-metadata-standards>. Acesso em: 05 out. 2019.

HIGGINS, S. The DCC Curation Lifecycle Model. **The International Journal of Digital Curation**, Edinburgh, n. 1, v. 3, 2008. Disponível em: <http://dx.doi.org/10.2218/ijdc.v3i1.48>. Acesso em: 30 jun. 2019.

HIGGINS, S. Digital curation: the emergence of a new discipline. **The International Journal of Digital Curation**, Edinburgh, n. 2, v. 6, 2011. Disponível em: <http://dx.doi.org/10.2218/ijdc.v6i2.191>. Acesso em: 30 jun. 2019.

HIGGINS, S. Digital curation: the development of a discipline within information science. **Journal of Documentation**, v.74, n.6, p.1318-1338, 2018. Disponível em: https://pure.aber.ac.uk/portal/files/26406315/Digital_curation_discipline_proof_final.pdf. Acesso em: 30 jun. 2019.

HIGGINS, S; SEMPLE, N. **OAIS Five-year review: Recommendations for update**, 2006. Disponível em: <http://hdl.handle.net/1842/3352>. Acesso em: 30 jun. 2019.

HJØRLAND, B. Information science and its core concepts: levels of disagreement. *In: IBEKWE-SANJUAN, F. DOUSA, T. (Ed.). Theories of information, communication and knowledge*. Dordrecht: Springer, 2014. p. 205-235.

HOUAISS, A. (Ed.). **Dicionário eletrônico Houaiss da língua portuguesa**. Instituto Antonio Houaiss. Rio de Janeiro, RJ: Editora Objetiva, 2009.

HYVÖNEN, E. **Publishing and using cultural heritage linked data on the semantic web**. EUA: Morgan & Claypool Publishers, 2012.

JORENTE, M. J. V.; SILVA, A. R.; PIMENTA, R. M. Cultura, memória e curadoria digital na plataforma SNIIC. **Liinc em revista**, Rio de Janeiro, v. 11, n. 1, 2015. Disponível em: <http://revista.ibict.br/liinc/article/view/3637/3101>. Acesso em: 30 jun. 2019.

JSON. **Introdução ao JSON**, [2020]. Disponível em: <https://www.json.org/json-pt.html>. Acesso em: 17 dez. 2019.

KIM, J. Growth and trends in digital curation research: the case of the international journal of digital curation. **American Society for Information Science and Technology**, Maryland, v. 51, n. 1, p. 1-4, 2014. Disponível em: <https://doi.org/10.1002/meet.2014.14505101074>. Acesso em: 17 dez. 2019.

LAVOIE, B.; DEMPSEY, L. Thirteen ways of looking at digital preservation. **D-Lib**, v. 10, n. 7/8, Jul./Ago. 2004. Disponível em: <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>. Acesso em: 30 jun. 2019.

LEE, C. A.; TIBBO, H. Where's the archivist in digital curation? Exploring the possibilities through a matrix of knowledge and skills. **Archivaria**, Ottawa, ON, v. 72, p. 123-168, 2011. Disponível em: <https://ils.unc.edu/callee/p123-lee.pdf>. Acesso em: 30 jun. 2019.

LEE, C. A.; TIBBO, H. R.; SCHAEFER, J. C. Defining what digital curators do and what they need to know: the DigCCurr project. *In: ACM/IEEE-CS joint conference on Digital libraries*, 7, ACM, 2007. p. 49-50. **Proceedings** [...]. Vancouver, BC, Canada: ACM/IEEE-CS joint conference on Digital libraries, 2007. Disponível em: 10.1145/1255175.1255183. Acesso em: 30 jun. 2019.

LIU, D.; BIKAKIS, A.; VLACHIDIS, A. Evaluation of semantic web ontologies for modelling art collections. **Springer International Publishing**: Nicosia, Cyprus, 2017. Disponível em: https://doi.org/10.1007/978-3-319-67162-8_34. Acesso em:

LORD, P. *et al.* From data deluge to data curation. *In: UK E-SCIENCE ALL HANDS MEETING*, 3, Londres, 2004. **Proceedings** [...]. Londres: National e-Science Centre, 2004. p. 371-375. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf>. Acesso em: 30 jun. 2019.

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Inf. & Soc.: Est.**, João Pessoa, v.21, n.1, p. 27-35, jan./abr. 2011. Disponível em:

<http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/7367/5596>. Acesso em: 30 jun. 2019.

MARCONDES, C. H. Representação e economia da informação. **Ci. Inf.**, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001. Disponível em: <http://www.scielo.br/pdf/ci/v30n1/a08v30n1>. Acesso em: 30 jun. 2019.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspect. Ciênc. Inf.** [online]. 2016, v.21, n.2, pp.61-83. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735/1748>. Acesso em: 30 jun. 2019.

MARCONDES, C. H. Relacionamentos culturalmente relevantes para interligar objetos do patrimônio digital na *web* usando tecnologias de dados interligados. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 19, 2018, Londrina, PR. **Anais [...]**. Londrina, PR: Universidade Estadual de Londrina, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/102416>. Acesso em: 14 jan. 2020.

MARON, N. L.; YUN, J.; PICKLE, S. **Sustaining our digital future**: institutional strategies for digital content. New York: Ithaka S+R, 2013. Disponível em: <https://sca.jiscinvolve.org/wp/files/2013/01/Sustaining-our-digital-future-FINAL-31.pdf>. Acesso em: 30 jun. 2019.

MÉNDEZ RODRÍGUEZ, E. **Metadados y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales**. Gijón: Trea, 2002. 429 p.

MICHAELIS. **Dicionário brasileiro da língua portuguesa Michaelis (online)**. [S.l.]: Editora Melhoramentos, 2020. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/>. 10 jan. 2020.

MORI, A.; CARVALHO, C. L. **Metadados no contexto da web semântica**. RT-INF_002-04. Goiânia: Instituto de Informática, Universidade Federal de Goiás, 2004. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_002-04.pdf. Acesso em: 30 jun. 2019.

NATIONAL RESEARCH COUNCIL. **Preparing the workforce for digital curation**. Washington, D.C: National Academies Press, 2015.

NOWACK, B. **The semantic web technology stack (not a piece of cake...)**. Linked Data Developer. [Online] 2009. Disponível em: <http://linkeddatadeveloper.com/Projects/Linked-Data/media/fig11.2.png>. Acesso em: 30 jun. 2019. 7 abr. 2016.

OLIVER, G.; HARVEY, R. **Digital curation**. London: Facet Publishing, 2016.

PENNOCK, M. Digital curation and the management of digital library cultural heritage resources. **Local Studies Librarian**, v. 25, n. 2, p. 3-7, 2006. Disponível em: http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lsl-curation_mep.pdf. Acesso em: 30 jun. 2019.

POOLE, A. H. The conceptual landscape of digital curation. **Journal of Documentation**, London, UK, v. 72, n. 5, p. 961-986, 2016. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JD-10-2015-0123/full/html>. Acesso em: 30 jun. 2019.

POMERANTZ, J. **Metadata**. Cambridge, Mass: MIT Press, 2015.

PRINCÍPIOS FAIR. **FAIR Principles**, [2019]. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 30 jun. 2019.

PURDAY, J. Think culture: Europeana.eu from concept to construction. **Bibliothek Forschung und Praxis**, v. 33, n. 2, p. 170-180, 2009. Disponível em: <https://doi.org/10.1515/bfup.2009.018>. Acesso em: 30 jun. 2019.

RAMALHO, R. A. S. **Web semântica**: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da ciência da informação. 2006. 120 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2006. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/ramalho_ras_me_mar.pdf. Acesso em: 30 jun. 2019.

RAMALHO, R. A. S.; OUCHI, M. T. Tecnologias semânticas: novas perspectivas para a representação de recursos informacionais. **Informação & Informação** v.16, n.3, p.60-75, 2011. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/9829/10643>. Acesso em: 10 jan. 2020.

RILEY, J. **Understanding metadata**. Bethesda: NISO Press, 2004. Disponível em: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>. Acesso em: 30 jun. 2019.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for? Baltimore: National Information Standards Organization (NISO), 2017. Disponível em: http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 30 jun. 2019.

RUSBRIDGE, C. *et al.* The digital curation centre: a vision for digital curation *In*: IEEE INTERNATIONAL SYMPOSIUM ON MASS STORAGE SYSTEMS AND TECHNOLOGY, Sardinia, Italy, 2005, p.31-41. **Proceedings** [...]. Sardinia, Italy: Mass Storage and Systems Technology Committee of the IEEE Computer Society, 2005. Disponível em: 10.1109/LGDI.2005.1612461. Acesso em: 30 jun. 2019.

SALES, R.; CAFÉ, L. Semelhanças e diferenças entre tesauros e ontologias. **Perspect. ciênc. Inf**, Belo Horizonte, v.14 n.1, Jan./Apr. 2009. Disponível em: <http://dx.doi.org/10.1590/S1413-99362009000100008>. Acesso em: 30 jun. 2019.

SALES, L. F.; SAYÃO, L. F. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da

informação, v. 17, n. esp. 2, p. 118-135, 2012. Disponível em: <https://doi.org/10.5007/1518-2924.2012v17nesp2p118>. Acesso em: 30 jun. 2019.

SANJAD, N.; BRANDÃO, C. R. F. A exposição como processo comunicativo na política curatorial. In: JULIÃO, L.; BITTENCOURT, J. N. (Coord.). **Cadernos de Diretrizes Museológicas**: mediação em museus: curadorias, exposições, ação educativa. Secretaria de Estado de Cultura de Minas Gerais, 2008. p. 35-43. Disponível em: http://www.cultura.mg.gov.br/files/museus/1caderno_diretrizes_museologicas_2.pdf. Acesso em: 30 jun. 2019.

SANTAREM SEGUNDO, J. E. **Recursos tecno-metodológicos para descrição e recuperação de informações na web**. 2004. 157 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2004. Disponível em: <http://repositorio.unesp.br/handle/11449/93618>. Acesso em: 30 jun. 2019.

SANTAREM SEGUNDO, J. E. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, PB, v. 8, n. 2, 2015. Disponível em: <http://www.brapci.inf.br/index.php/article/download/43737>. Acesso em: 30 jun. 2019.

SANTAREM SEGUNDO, J. E.; CONEGLIAN, C. S. Web semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. **Informação & Informação**, v.21, n. 2, p.217-244, 2016. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2016v21n2p217>. Acesso em: 10 jan. 2020.

SANTOS, T. N. C. **Curadoria digital**: o conceito no período de 2000 a 2013. 2014.165 f. Dissertação (Mestrado em Ciência da Informação). Programa de Pós-Graduação em Ciência da Informação da Faculdade de Ciência da Informação, Universidade de Brasília. Brasília, 2014. Disponível em: <http://repositorio.unb.br/handle/10482/17324>. Acesso em: 30 jun. 2019.

SANTOS, P. L. V. A. C.; ALVES, R. C. V. Metadados e web semântica para estruturação da Web 2.0 e Web 3.0. **DataGramZero**, Rio de Janeiro, v. 6, n. 10, 2009. Disponível em: <http://www.brapci.ufpr.br/brapci/index.php/article/download/52958>. Acesso em: 30 jun. 2019.

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**, v. 42, n. 2, jan. 2013. Disponível em: <http://revista.ibict.br/ciinf/article/view/1382/1560>. Acesso em: 10 jan, 2020.

SANTOS, P. L. V. A. C. SIMIONATO, A. C.; ARAKAKI, F. A. definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Informação & Informação**, Londrina, v. 19, n. 1, 2014, p. 146-163. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2014v19n1p146>. Acesso em: 30 jun. 2019.

SAYÃO, L. F. Metadados de preservação: informações para a gestão da preservação de objetos digitais. In: SILVA, M. C. S. M. **Segurança de acervos culturais**. Rio de Janeiro: Museu de Astronomia e Ciências Afins, 2012. 200p. Disponível em:

http://200.156.20.14/images/pdf/publicacoes_do_mast/seguran%C3%A7a_de_acervos_culturais.pdf#page=110. Acesso em: 30 jun. 2019.

SAYÃO, L. F. Digitalização de acervos culturais: reuso, curadoria e preservação. *In: SEMINÁRIO SERVIÇOS DE INFORMAÇÃO EM MUSEUS*, 4, 2016, São Paulo. **Anais** [...]. São Paulo: [s.n.], 2016, p. 47-61. Disponível em: https://www.researchgate.net/publication/319403030_Digitalizacao_de_acervos_culturais_reuso_curadoria_e_preservacao. Acesso em: 12 dez. 2019.

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133-148, 2008. Disponível em: <http://dx.doi.org/10.1590/S0103-37862008000200002>. Acesso em: 30 jun. 2019.

SAYÃO L. F. SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade: Estudos**, v.22, n.3, p.179-191, 2012. Disponível em: <https://search.proquest.com/openview/82ecc12d22135fd5a2510ac8fdeec623/1?pq-origsite=gscholar&cbl=2030753>. Acesso em: 30 jun. 2019.

SAYÃO, L. F. SALES, L. F. Curadoria digital e dados de pesquisa. **Atoz**: novas práticas em informação e conhecimento, v.5, n. 2, p.67-71, 2016. Disponível em: <http://dx.doi.org/10.5380/atoz.v5i2.49708>. Acesso em: 30 jun. 2019.

SENSO, J. A.; ROSA PIÑERO, A. El concepto de metadato: algo más que descripción de recursos electrónicos. **Ciência da Informação** v.32, n. 2, 2003. Disponível em: <http://www.scielo.br/pdf/ci/v32n2/17038.pdf/>. Acesso em: 30 jun. 2019.

SIEBRA, S. A.; BORBA, V. R.; MIRANDA, M. K. F. O. Curadoria digital: um termo interdisciplinar. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 17, Salvador, 2016. **Anais**[...]. Salvador: Universidade Federal da Bahia, 2016. Disponível em: <http://www.brapci.inf.br/index.php/article/download/48573>. Acesso em: 30 jun. 2019.

SIEBRA, S. A. *et al.* Curadoria digital: além da questão da preservação digital. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 14, Florianópolis, 2013. **Atas eletrônicas**[...]. Florianópolis: Universidade Federal de Santa Catarina, 2013. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2478>. Acesso em: 30 jun. 2019.

SILVA, F. M. O. **Curadoria digital**: recomendações para acervos de objetos culturais digitais. 2017. 227 f. Dissertação (Mestrado em Ciência da Informação). Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, Recife, PE, 2007. Disponível em: <https://repositorio.ufpe.br/handle/123456789/27700>. Acesso em: 30 jun. 2019.

SILVA, S. D; LOUREIRO, J. M. M. Gênese e singularidades nos processos curatoriais nos espaços de história natural: dos gabinetes aos museus como espaços discursivos da ciência e

da “idéia de nação”. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14, 2013, Florianópolis. **Anais**[...]. Florianópolis: UFSC, 2013. p. 1 - 18. Disponível em:

<http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2549>. Acesso em: 30 jun. 2019.

SIMIONATO, A. C. **Modelagem conceitual DILAM**: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital. 2015. 200 f. Tese (doutorado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências, 2015. Disponível em: <http://hdl.handle.net/11449/123318>. Acesso em: 30 jun. 2019.

SOUZA, R. R.; ALVARENGA, L. A Web semântica e suas contribuições para a ciência da informação. **Ci. Inf.**, Brasília, v. 33, n. 1, p. 132-141, jan./abril 2004. Disponível em: <http://www.scielo.br/pdf/ci/v33n1/v33n1a16>. Acesso em: 30 jun. 2019.

THOMAZ, K. P.; SOARES, A. J. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero**, Rio de Janeiro, v. 5, n. 1, p. 1-17, 2004. Disponível em: http://www.brapci.inf.br/_repositorio/2010/01/pdf_99df3bbc0d_0007616.pdf. Acesso em: 30 jun. 2019.

TIBBO, H., HANK, C. LEE, C.A. Challenges, curricula, and competencies: researcher and practitioner perspectives for informing the development of a digital curation curriculum. *In*: SOCIETY FOR IMAGING SCIENCE AND TECHNOLOGY ARCHIVING CONFERENCE, Bern, Switzerland, 2008. **Proceedings** [...]. Bern, Switzerland: Society for Imaging Science and Technology, 2008. Disponível em: <https://www.ingentaconnect.com/content/ist/ac/2008/00002008/00000001/art00048>. Acesso em: 30 jun. 2019.

TRIQUES, M. L. **Metadados na curadoria**: uma perspectiva na Ciência da Informação, 2017. 67 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia e Ciência da Informação) - Departamento de Ciência da informação, Universidade Federal de São Carlos, São Carlos, 2017.

UK DATA ARCHIVE. **We are a critical part of the UK’s research infrastructure**, 2002-2019. Disponível em: <https://www.data-archive.ac.uk/>. Acesso em: 31 jun. 2019. 30 jun. 2019.

WEBER, C. Curadoria digital de dados científicos: pelo viés de um periódico. **P2P & inov.** Rio de Janeiro, v. 3 n. 1, set./mar. 2017. Disponível em: <https://doi.org/10.21721/p2p.2016v3n1.p130-147>. Acesso em: 30 jun. 2019.

WEIDNER, A. J.; ALEMNEH, D. G. Workflow tools for digital curation. **Code4Lib Journal**, n. 20, 2013. Disponível em: <https://journal.code4lib.org/articles/8419>. Acesso em: 30 jun. 2019.

WINER, D.; ROCHA, I. E. Europeana: um projeto de digitalização e democratização do patrimônio cultural europeu. **Patrimônio e Memória**, 2013. Disponível em: <http://pem.assis.unesp.br/index.php/pem/article/view/327>. Acesso em: 10 jan. 2020.

WORLD WIDE WEB CONSORTIUM. **Naming and addressing: URIs, URLs, ...**, 2006. Disponível em: <https://www.w3.org/Addressing/>. Acesso em: 17 dez. 2019.

WORLD WIDE WEB CONSORTIUM. **Turtle: terse RDF Triple Language**, 2011. Disponível em: <https://www.w3.org/TeamSubmission/turtle/#sec-tutorial>. Acesso em: 17 dez. 2019.

WORLD WIDE WEB CONSORTIUM. **RIF overview**. 2.ed., 2013a. Disponível em: <https://www.w3.org/TR/rif-overview/>. Acesso em: 30 jun. 2019.

WORLD WIDE WEB CONSORTIUM. **PROV-O: a ontologia PROV**, 2013b. Disponível em: <https://www.w3.org/TR/prov-o/>. Acesso em: 10 jan. 2020.

WORLD WIDE WEB CONSORTIUM. **RDF 1.1 Primer**, 2014. Disponível em: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>. Acesso em: 30 jun. 2019.

WORLD WIDE WEB CONSORTIUM. **Linked data**, 2015. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 30 jun. 2019.

WORLD WIDE WEB CONSORTIUM. **Extensible Markup Language (XML)**, 2016. Disponível em: <https://www.w3.org/XML/>. Acesso em: 30 jun. 2019.

WORLD WIDE WEB CONSORTIUM. **Data on the web best practices**, 2017. <https://www.w3.org/TR/dwbp/>. Acesso em: 10 jan. 2020.

YAKEL, E. Digital Curation. **OCLC Systems & Services**, v.23, n.4, 2007, p. 335-340. Disponível em: <https://doi.org/10.1108/10650750710831466>. Acesso em: 30 jun. 2019.

YAMAOKA, E. J. Ontologia para mapeamento da dependência tecnológica de objetos digitais no contexto da curadoria e preservação digital. **AtoZ**, Curitiba, v. 1, n. 2, p. 65-78, jan./dez. 2012. Disponível em: <http://revistas.ufpr.br/atoz/article/view/41313/25240>. Acesso em: 30 jun. 2019.

ZENG, M. L.; QIN, J. **Metadata**. Chicago, IL: ALA Neal-Schuman, 2008.

ZENG, M. L.; QIN, J. **Metadata**. 2.ed. Chicago, IL: ALA Neal-Schuman, 2016.