UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO MESQUITA FILHO"
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

FELIPE AUGUSTO ARAKAKI

METADADOS ADMINISTRATIVOS E A PROVENIÊNCIA DOS DADOS: MODELO BASEADO NA FAMÍLIA PROV

Marília/SP
2019

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO MESQUITA FILHO"
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

FELIPE AUGUSTO ARAKAKI
ORCID: http://orcid.org/0000-0002-3983-2563
Lattes: http://lattes.cnpq.br/5324289839207169

METADADOS ADMINISTRATIVOS E A PROVENIÊNCIA DOS DADOS: MODELO BASEADO NA FAMÍLIA PROV

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista (UNESP), como requisito para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento

Linha de Pesquisa: Informação e Tecnologia

Orientadora: Profa. Dra. Plácida Leopoldina Ventura Amorim da Costa Santos

Marília/SP
2019

A659m

Arakaki, Felipe Augusto
Metadados administrativos e a proveniência dos dados : modelo baseado na família PROV / Felipe Augusto Arakaki. -- , 2019
139 f. : il.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília
Orientadora: Plácida Leopoldina Ventura Amorim da Costa Santos

1. Metadados. 2. Catalogação. 3. Tecnologia da informação. 4. Família PROV. 5. Proveniência dos dados. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

FELIPE AUGUSTO ARAKAKI

METADADOS ADMINISTRATIVOS E A PROVENIÊNCIA DOS DADOS: MODELO BASEADO NA FAMÍLIA PROV

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista (UNESP), como requisito para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento.
Linha de Pesquisa: Informação e Tecnologia
Data da defesa: 11 de janeiro de 2019.
Local: Faculdade de Filosofia e Ciências, UNESP - Campus de Marília

BANCA EXAMINADORA

Presidente e Orientadora: Profa. Dra. Plácida Leopoldina Ventura Amorim da Costa Santos
Professora no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP Campus de Marília

Membro Titular: Profa. Dra. Silvana Aparecida Borsetti Gregorio Vidotti
Professora no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP Campus de Marília

Membro Titular: Profa. Dra.Rachel Cristina Vesu Alves
Professora no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP Campus de Marília

Membro Titular Externo: Prof. Dr. Fabiano Ferreira de Castro
Professor no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos, UFSCar

Membro Titular Externo: Dr. Fabrício Silva Assumpção
Bibliotecário na Universidade Federal do Paraná, UFPR Setor Litoral

À Carolina pelo carinho, amor e apoio incondicional em todos os momentos.

AGRADECIMENTOS

À Carolina pelo apoio incondicional, pela paciência, pelo incentivo e sempre esteve ao meu lado em todos os momentos!

Aos meus pais Carmo e Luiza, meus irmãos Cristiane, Fernando, Letícia, Noemia e Ricardo pelo incentivo. À toda minha família, tios, tias, primos, primas, cunhados, sobrinhas.

À dona Roseli, Arnaldo e Anselmo pelo apoio e carinho.

Às professoras Plácida, Rachel e Silvana por todo o incentivo no decorrer desses anos, por sua compreensão, amizade, carisma, pela paciência, ensinamentos, confiança e serei eternamente grato por sempre acreditar em mim e principalmente por me guiar e direcionar para área acadêmica.

À minha banca, professores Rachel, Silvana, Fabiano e Fabrício, pelas contribuições para o desenvolvimento do trabalho e professora Ana Alice pelas contribuições para o desenvolvimento do trabalho, principalmente na banca de qualificação do trabalho.

Aos meus estimados professores Brígida, Edberto, Eduardo, Fernando Vechiato, Henry, José Augusto, Juan Pastor, Leonardo Botega, Maria Cláudia, Natália, Paula Amorim, Plácida, Rachel, Ricardo, Rosângela, Silvana Drumond, Silvana Vidotti, Walter, Zaira e aos professores do Programa de Pós-Graduação da Ciência da Informação da Unesp pela dedicação, ensinamentos e que sempre serão minha fonte inspiração.

Aos colegas do repositório da UNESP Juliano, Ana Paula e Luiza que me deram apoio para sempre continuar e Luiza, Bruna, Luiz e Monique pela amizade.

Aos colegas do IFSP de Presidente Epitácio e Itapetininga, em especial da biblioteca Claudinei, Fabiana e Vanderlei, pelo incentivo e amizade. À Joelma pela parceria dos projetos da biblioteca e amizade. Ao Vinícius, Vinícius Black, Filippo, Willian, Ricardo, Maycon, Jeferson, Diego, Zé Hélio, Joselita, Aline, Audrie, Eliane, Josy, Mitsuko. À Marcia Jani e Zé Guilherme pelo apoio para seguir no doutorado. Aos envolvidos nas parcerias e projetos da biblioteca, servidores e bolsistas e demais alunos do IFSP-PEP. Ao Willian, Paulinho e Carol.

Aos colegas da rede de bibliotecas do IFSP e das comissões do Repositório Institucional do IFSP, Editora do IFSP e do mapeamento dos processos das bibliotecas, em especial à Angela, Luis e Andréia pela confiança e parceria.

Aos colegas do GP-NTI, em especial, Caio, Sandra, Jacquelin, Ana Maria, Edgar, Luiza, Fernando, Mariana, Laís, Luciana, Liliana, Diana, Bete, Manu que me acompanharam nessa

jornada.

Aos funcionários da UNESP Marília que sempre me ajudaram nos momentos em que precisavam, em especial da biblioteca Neuza, Vania, Eliza, Bete, Janaína, Satie, André, Telma e da Seção Técnica de Pós-Graduação.

Aos amigos do grupo de Taiko Yuuyake Wadaiko, em especial Keiji, Black, Giovani, Bea, Vitor, Igor, Estéfani e Kátia!

À UNESP e o IFSP pelos apoios para realização da pesquisa.

A todos que contribuíram diretamente e indiretamente no desenvolvimento deste trabalho.

Obrigado por tudo!

RESUMO

O catálogo é um ambiente pelo qual os usuários podem encontrar, identificar, selecionar e navegar para obter um recurso informacional. Seu desenvolvimento sempre esteve atrelado ao uso das tecnologias disponíveis, com o objetivo de aperfeiçoar e agilizar o processo de busca, localização, acesso e de recuperação. A base para esse instrumento é a construção de formas de representação realizadas por meio dos metadados. Entretanto, com a expansão e popularização da publicação de dados na Web, são necessários sistemas cada vez mais interoperáveis e alguns problemas ainda não foram solucionados como a identificação da origem, registros de ações, entre outras informações no domínio bibliográfico, principalmente no que diz respeito aos padrões de metadados, a abertura dos catálogos e repositórios digitais para o reaproveitamento de dados de bibliográficos. Nesse contexto a questão central desta pesquisa foi: qual a função dos metadados de proveniência nos registros bibliográficos em ambientes digitais? A partir da questão norteadora desta tese, considera-se que a catalogação pode auxiliar na construção de representações a partir dos metadados administrativos e de proveniência para permitir a confiabilidade e integridade dos dados de bibliotecas, e principalmente, a catalogação influencia diretamente na construção de descrições dos recursos informacionais em catálogos e repositórios digitais persistindo as informações nos registros bibliográficos. Nesse contexto, a hipótese da tese configura-se que o modelo PROV-O, modelo baseado na família PROV, pode ser aplicado ao domínio bibliográfico para a representação da proveniência em registros bibliográficos, permitindo a descrição da origem, das ações e dos envolvidos na construção e alteração de registros em ambientes digitais. Dessa forma, a tese consiste em que os registros bibliográficos necessitam de metadados referentes à proveniência dos dados para a preservação da integridade e da persistência, como forma na garantia da confiabilidade das informações em ambientes digitais. A presente proposta parte da necessidade de reutilizar dados em catálogos de bibliotecas e repositórios digitais, independentemente do padrão de metadados utilizado. Destaca-se a importância da proveniência dos dados na perspectiva do reuso dos dados em catálogos e repositórios digitais por meio de um modelo de dados que seja interoperável. Para tanto, o objetivo geral é analisar a viabilidade da aplicação do modelo PROV-O para a representação da proveniência em registros bibliográficos de ambientes digitais. Caracteriza-se por uma pesquisa qualitativa, em razão da análise que busca entender o relacionamento entre os metadados administrativos, a proveniência dos dados no domínio bibliográfico. Como resultados apresentou o crosswalk entre o PROV-O, MARC21, Dublin Core, PREMIS, BIBFRAME e Schema.org para verificar a compatibilidade dos padrões com a questão da proveniência. O estudo sobre a proveniência revelou que a temática no Brasil é incipiente e carece de pesquisas teóricas e iniciativas de cunho prático/profissional. Tal conclusão, revela a necessidade de um cuidado especial no planejamento do sistema, na definição dos metadados que irão compor o registro bibliográfico do recurso informacional. Isso irá refletir na definição de quais informações são necessárias aos usuários, para o acesso e a visualização e, por último, e não menos importante, na definição de quais são informações necessárias para gestão e curadoria do sistema.

Palavras-chave: Proveniência dos dados; Família PROV; Metadados; Catalogação; Crosswalk; Domínio bibliográfico

ABSTRACT

The catalog is an environment by which users can find, identify, select and navigate to obtain an informational resource. Its development has always been linked to the use of available technologies, with the objective of improving and streamlining the search, localization, access and retrieval process. The basis for this instrument is the construction of forms of representation performed through the metadata. However, with the expansion and popularization of data publication on the Web, increasingly interoperable systems are needed and some problems have not yet been solved, such as origin identification, action records, among other information in the bibliographic domain, especially with respect to to the metadata standards, the opening of catalogs and digital repositories for the reuse of bibliographic data. In this context the central question of this research was: what is the function of provenance metadata in bibliographic records in digital environments? From the guiding question of this thesis, it is considered that the cataloging can help in the construction of representations from the administrative and provenance metadata to allow the reliability and integrity of the data of libraries, and mainly, the cataloging directly influences the construction of descriptions of the information resources in catalogs and digital repositories, persisting the information in the bibliographic records. In this context, the hypothesis of the thesis is that the PROV-O model, based on the PROV family, can be applied to the bibliographic domain for the representation of provenance in bibliographic records, allowing the description of origin, actions and involved in the construction and alteration of records in digital environments. Thus, the thesis consists in that the bibliographic records need metadata referring to the provenience of the data for the preservation of the integrity and the persistence, as a way to guarantee the reliability of the information in digital environments. The present proposal starts from the need to reuse data in catalogs of libraries and digital repositories, regardless of the metadata standard used. The importance of the provenance of the data from the perspective of the reuse of the data in catalogs and digital repositories through an interoperable data model is highlighted. For this, the general objective is to analyze the feasibility of the application of the PROV-O model for the representation of provenance in bibliographic records of digital environments. It is characterized by a qualitative research, due to the analysis that seeks to understand the relationship between administrative metadata, the provenience of the data in the bibliographic domain. As results presented the crosswalk between PROV-O, MARC21, Dublin Core, PREMIS, BIBFRAME and Schema.org to check the compatibility of standards with the question of provenance. The study on provenance revealed that the thematic in Brazil is incipient and lacks theoretical research and initiatives of a practical / professional nature. This conclusion reveals the need for special care in the planning of the system, in the definition of the metadata that will compose the bibliographic record of the information resource. This will reflect in defining what information is required for users, for access and viewing and, last but not least, in defining what information is needed for system management and curation.

Keywords: Data Provenance; PROV family; Metadata; Cataloguing; Crosswalk; Bibliographic domain

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

AACR	*Anglo-American Cataloguing Rules*
AC	*Administrative Component*
API	*Application Programming Interface*
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
BENANCIB	Repositório Questões em Rede – Coleções
BIBFRAME	*Bibliographic Framework Initiative*
BRAPCI	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
DCAM	*Dublin Core Abstract Model*
DCAP	*Dublin Core Application Profiles*
DCMI	*Dublin Core Metadata Initiative*
EDS	*EBSCO Discovery Service*
EXIF	*Exchangeable image file format*
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FRBR	*Functional Requirements for Bibliographic Records*
GIF	*Graphics Interchange Format*
HTML	*Hypertext Markup Language*
IFLA	*International Federation of Library Associations and Institutions*
IFSP	Instituto Federal de Educação, Ciência e Tecnologia São Paulo
IPAW	*International Provenance and Annotation Workshop*
ISBD	*International Standard Bibliographic Description*
ISO	*International Organization for Standardization*
JOSON-LD	*JavaScript Object Notation for Linked Data*
JPG	*Joint Photographic Experts Group*
JCR	*Journal Citation Reports*
LC	*Library of Congress*
LISA	*Library Information Science Abstratcs*
LISTA	*Library, Information Science and Technology Abstracts*
MARC	*Machine Readable Cataloging*
METS	*Metadata Encoding & Transmission Standard*
MXF	*Material eXchange Format*
NSDL	*National Science Digital Library*
OAI-ORE	*Open Archives Initiative Object Reuse and Exchange*
OAI-PMH	*Open Archives Initiative Protocol for Metadata Harvesting*
OAIS	*Open Archival Information System*
OCLC	*Online Computer Library Center*
OPAC	*Online Public Access Catalogue*
OPM	*Open Provenance Model*
OWL	*Web Ontology Language*
PAV	*Provenance, Authoring and Versioning*
PDF	*Portable Document Format*
PREMIS	*PREservation Metadata: Implementation Strategies*
PROV	*Provenance*
PROV-CONSTRAINTS	*Constraints of the PROV Data Model*

PROV-DM	*The PROV Data Model*
PROV-N	*The Provenance Notation*
PROV-O	*The PROV Ontology*
RDF	*Resource Description Framework*
RDFa	*Resource Description Framework in Attributes*
RDs	Repositórios Digitais
RLG	*Research Libraries Group*
SciELO	*Scientific Electronic Library Online*
SJR	*Scientific Journal Rankings*
SWfMS	*Scientific Workflow Management Systems*
UML	*Unified Modeling Language*
URI	*Uniform Resource Identifier*
URL	*Uniform Resource Locator*
W3C	*World Wide Web Consortium*
XML	*Extensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO

O catálogo é um ambiente que permite aos usuários encontrar, identificar, selecionar e navegar para obter um recurso informacional. Para o seu desenvolvimento, sempre esteve atrelado ao uso das tecnologias disponíveis, com o objetivo de aperfeiçoar e agilizar o processo de busca, de localização, de acesso e de recuperação. Sendo que a principal base para estruturação do catálogo é a construção de formas de representação realizadas por meio dos metadados. Os metadados são atributos referenciais de um recurso e desempenham um papel fundamental para representação, gestão, recuperação e interoperabilidade dos recursos informacionais.

Com o movimento de publicação de dados, torna-se cada vez mais importante estruturar os ambientes digitais para que os usuários possam localizar e recuperar informações desejadas. Com isso, ampliam-se as possibilidades de reuso de dados como alternativa de evitar o retrabalho de descrição um mesmo recurso informacional. Nesse contexto, Santos e Sant'Ana (2013, p. 200) discutem que a Ciência da Informação como área do conhecimento focada nas "[...] metodologias e nos instrumentos desenvolvidos ao longo do tempo para armazenar, descrever, recuperar, preservar, disseminar e compartilhar as experiências humanas."

Como forma de solucionar as questões de representação e de organização da informação, a área da Ciência da Informação proporciona estudos teóricos e aplicados sobre o comportamento informacional, os fluxos e os usos da informação com intuito do armazenamento, da recuperação e da disseminação (BORKO, 1968) dos recursos informacionais. Conforme apontado por Santos e Sant'Ana (2013, p. 200),

> A ciência da informação refere-se à atividade direcionada à pesquisa de princípios e métodos que são partes da análise, do projeto e da evolução dos sistemas de informação. Nesses sistemas, os elementos constituintes são o ambiente, as pessoas, os recursos informacionais, as tecnologias e os procedimentos. Eles sustentam a capacidade para a busca de soluções e tomada de decisões como parte da vida diária, envolvendo a manipulação de dados, o acesso à informação e a apropriação do conhecimento.

Saracevic (1996), por sua vez, explica que a evolução da Ciência da Informação possui basicamente três características. A primeira que a Ciência da Informação é interdisciplinar por natureza, a segunda, é que está ligada às tecnologias da informação e a terceira que

desempenha um importante papel na dimensão social e humana. No contexto desta pesquisa, a interdisciplinaridade está principalmente na relação entre a Ciência da Informação e a Ciência da Computação no que diz respeito aos ambientes digitais e na estrutura de ambientes de representações. Já a questão da dimensão social e humana está relacionada às contribuições para construção de registros com maior integridade e ampliação das possibilidades navegacionais em catálogos, possibilitando assim, a identificação de informações de origem dos dados acessados.

O compartilhamento e o reuso dos dados (registros) é tratado pela Biblioteconomia, em especial, pela catalogação no que diz respeito à modelagem, a construção e à alimentação de catálogos. Nesse sentido, ao analisar o movimento de dados abertos na *Web*, inclui-se os procedimentos para a abertura dos catálogos para busca, recuperação e integração com outras fontes de informação. É nesse contexto que os metadados administrativos têm papel fundamental para garantir a proveniência do registro e as devidas atribuições às agências publicadores de dados.

1.1 Problema de pesquisa

Com a expansão e popularização da publicação de dados na *Web*, são necessários sistemas cada vez mais interoperáveis. Entretanto, alguns problemas ainda não foram solucionados como a identificação da origem, registros de ações, entre outras informações no domínio bibliográfico. Principalmente no que diz respeito aos padrões de metadados, a abertura dos ambientes digitais para o reaproveitamento de dados de bibliográficos.

A escolha do esquema de metadados adotado e a garantia de informações referentes a quem está publicando os dados e quando. Essas informações são exemplos de metadados administrativos que devem ser persistidos no registro bibliográfico para facilitar na identificação da proveniência.

Ressalta-se que diversos padrões têm sido apresentados como alternativas para construção de representações, possibilitando ampliar a recuperação e o acesso às informações. Entretanto, as informações sobre a origem podem se perder com descrições inadequadas.

O Provenance (PROV) possui um conjunto de documentos genérico publicado pelo

World Wide Web Consortium (W3C) em 2013, para representar especificamente informações de proveniência em qualquer domínio e pouco foi discutido sua aplicação no domínio bibliográfico.

Por outro lado, o *Dublin Core* é um padrão de metadados reconhecido e aceito internacionalmente, utilizado principalmente no contexto de repositórios digitais. Para maximizar a representação e completar descrições utilizando o *Dublin Core*, Garijo e Eckert (2013) apresentaram um mapeamento do *Dublin Core* com metadados de proveniência.

Em alternativa, a *Library of Congress* (LC) que gerencia diversos padrões de metadados utilizados no mundo todo, vem assumindo um papel importante em ligar dados de bibliotecas baseando no *Bibliographic Framework Initiative* (BIBFRAME) como substituto do *Machine Readable Cataloging* (MARC) 21 e que pouco têm-se discutido a questão da proveniência dos dados, além do padrão para preservação, o *PREservation Metadata: Implementation Strategies* (PREMIS).

Em contrapartida, o *Schema.org* é utilizado para representar recursos informacionais no catálogo do *WorldCat* administrado pela *Online Computer Library Center* (OCLC) mas ainda não há um modelo que garanta a proveniência dos dados compartilhados.

Dessa forma, a reutilização dos dados sem a devida preocupação da descrição, pode causar dificuldade da identificação do criador de determinados registros, e também, o seu pertencimento e proveniência. Nesse contexto a questão central desta pesquisa é: **qual a função dos metadados de proveniência nos registros bibliográficos em ambientes digitais?**

1.2 Hipótese e tese

Considerando a questão norteadora desta tese, afirma-se que a catalogação influencia diretamente na construção de descrições dos recursos informacionais em ambientes digitais persistindo os registros bibliográficos, permitindo a confiabilidade e integridade dos dados de bibliotecas. Nesse contexto, a **hipótese da tese** configura-se que o modelo PROV-O pode ser aplicado ao domínio bibliográfico para a representação da proveniência em registros bibliográficos, permitindo a descrição da origem, das ações e dos envolvidos na construção e alteração de registros em ambientes digitais.

Nesse contexto, a **tese** consiste em que os registros bibliográficos necessitam de metadados referentes à proveniência dos dados para a preservação da integridade e da

persistência, como forma na garantia da confiabilidade das informações em ambientes digitais.

1.3 Justificativa e relevância da tese

Esta pesquisa corrobora com o campo científico no desenvolvimento das temáticas em questão, como a proveniência, metadados e a proveniência em padrões de metadados. Observou-se a escassez de estudos teóricos sobre os metadados administrativos, em especial a proveniência dos dados no Brasil. Assim sendo, a discussão proposta oferece subsídios na fundamentação teórica também em áreas correlatas como a Ciência da Computação, corroborando com o conhecimento científico no processo de reutilização e de interoperabilidade dos dados.

Além disso, pretende-se dar continuidade ao trabalho de Iniciação Científica financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que foi estudado a evolução do *Dublin Core* e suas contribuições no domínio bibliográfico. Também a pesquisa realizada, durante o mestrado, foi possível verificar a aplicação do *Linked Data* em catálogos e sua relação com os padrões de metadados *Dublin Core*, BIBFRAME e Schema.org.

Igualmente, se registra o fato de que a presente pesquisa coadunar-se com a tradição de pesquisa no tocante aos estudos do Programa de Pós-Graduação em Ciência da Informação (PPGCI)/Unesp - Marília, em especial da linha de pesquisa “Informação e Tecnologia” que

> Reflete sobre as questões apresentadas pelos ambientes informacionais digitais para a construção do conhecimento e da experimentação em torno de novas formas de acesso; de **organização**; de **representação**, de recuperação; de políticas; e de processamento de dados e de informação para a otimização e a personalização de processos e de sistemas informacionais **em distintas ambiências no campo de conhecimento da Ciência da Informação**. (PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - UNESP, 2018, não paginado, grifo nosso).

Dessa forma, há o desejo de realizar o aprofundamento teórico nas questões de representação e de organização de dados, em especial, relativos aos metadados administrativos e à proveniência dos dados, pela tradição da pesquisa da linha e dos trabalhos desenvolvidos durante a graduação e o mestrado. No âmbito profissional, pelo desenvolvimento dessas temáticas inicialmente no Repositório Institucional UNESP e

posteriormente como bibliotecário e catalogador do O Instituto Federal de Educação, Ciência e Tecnologia São Paulo (IFSP) e membro da comissão de implementação do Repositório Institucional do IFSP.

A discussão proposta contribui para prática profissional, pois, a identificação clara da proveniência dos dados a partir dos metadados administrativos é fundamental para promover a credibilidade das informações reutilizadas em ambientes digitais.

Já na área social, o estudo irá proporcionar discussões sobre a aplicação da proveniência dos dados e que poderão ser ampliadas em outros contextos, principalmente no ambiente digital. A concretização da proposta facilitará a identificação das informações, possibilitando a localização de diversos outros recursos relacionados ao que se está visualizando. Logo, o usuário poderá otimizar a forma de navegar no catálogo, podendo escolher um determinado recurso informacional, suas derivações e recursos relacionados.

1.4 Teoria de base

Com os apontamentos sobre a pesquisa elucidados, destaca-se que a teoria de base possibilita a discussão e o aprofundamento teórico em metadados administrativos conforme abordado por Méndez Rodríguez (2002); Haynes (2004, 2018); Liu (2007); Zeng e Qin (2008, 2016); Alves (2010); Miller (2011); Alves e Santos (2013); Pomerantz (2015); Baca (2016); Riley (2017). Os metadados administrativos provêm informações sobre a origem e a manutenção do recurso, entretanto, não há consenso na literatura se os metadados de proveniência estão contemplados ou não na categoria de metadados administrativos.

Nesse contexto, para alguns autores (BACA, 2016; CHOWDHURY; CHOWDHURY, 2007; LIU, 2007; MILLER, 2011; POMERANTZ, 2015; TAYLOR; JOUDREY, 2009; ZENG; QIN, 2008) os metadados administrativos podem englobar como subcategorias os metadados de preservação, metadados estruturais, metadados técnicos e metadados de uso.

Para Glushko (2013) um recurso necessita da garantia da permanência dos dados ao longo do tempo definida como persistência, em consequência terá a efetividade, a autenticidade e por fim a proveniência das informações. Glushko (2013) aponta que há duas formas de **persistência**, a primeira está relacionada aos identificadores persistentes e a segunda está relacionada com a persistência dos recursos que envolve a preservação do

mesmo. A **efetividade** está relacionada ao tempo de vida do recurso, ou seja, o período em que esse recurso é válido. A **autenticidade** está relacionada à certificação que um documento possui em relação ao seu original. Já a **proveniência** está relacionada a custódia do recurso, ou seja, a quem ele pertence. Para Moreau e Missier (2013) a proveniência são informações referentes às entidades, às atividades e s pessoas envolvidas na produção de um dado ou coisa, que pode ser usado para formar avaliações sobre a sua qualidade, segurança ou confiabilidade. Em paralelo a esses pilares, os metadados são de extrema importância para registrar essas informações e suas alterações ao longo do tempo.

Já os trabalhos correlatos, pode-se citar o trabalho de Kumar, Ujjal e Utpal (2013) que fazem a adaptação do Formato MARC 21 para Dados Bibliográficos baseado nas triplas *Resource Description Framework* (RDF) e nos princípios do *Linked Data*, assim os autores fazem a conversão a partir da proposta do *MARCOnt* e apresentaram algumas informações para garantir a proveniência das informações de forma automática. Li e Sugimoto (2014, 2017, 2018) apresentaram uma proposta de um modelo de proveniência para perfis de aplicação e discussões sobre a proveniência e o PREMIS. Os estudos de Eckert (2012) que discute a questão da proveniência na *Europeana Data Model* e de Garijo e Eckert (2013) no mapeamento do PROV para *Dublin Core*.

Há discussões também, sobre os metadados administrativos e a proveniência dos dados no contexto da *Web* Semântica e *Linked Open Data* com Moreau et al. (2011) e Jaques et al. (2012).

No contexto das bibliotecas, Baker et al. (2011) esclarecem que 'Dados de Biblioteca' são definidos a qualquer tipo de informação digital que descrevem recursos ou ajudam a sua descoberta produzidos ou gerenciados por biblioteca. Entretanto, vale destacar que em muitos casos dados e metadados são tratados como sinônimos, pois Jeffery et al. (2013, não paginado, tradução nossa) explicam que

> [...] para o pesquisador, o registro da biblioteca são metadados para descobrir um livro ou artigo de interesse. Para o bibliotecário, o registro pode ser utilizado como dados para analisar a completude relativa das coleções por assunto, por editora, por ano etc.

Nesse sentido, ressalta-se que neste trabalho os conceitos de dados e metadados serão adotados como sinônimos, seguindo a aderência de Santos e Sant'Ana (2013) para dados e Alves (2010) para metadados e a concepção de Jeffrery (2013).

A representação da informação por meio dos metadados é fundamental para estruturação de dados em ambientes digitais. Assim, os metadados são base para possibilitar a descrição, recuperação, gerenciamento, interoperabilidade e a autenticidade digital de um recurso. (CHOWDHURY; CHOWDHURY, 2007; HAYNES, 2004).

Assim sendo, a tese é fundamentada por autores da catalogação como Mey (1995), Mey e Silveira (2009), Santos (2008), Alves (2010), Castro e Santos (2014), Simionato (2015), Assumpção (2018) no intuito de "[...] construir as formas de representação para alimentação de catálogos a partir da descrição padronizada de recursos informacionais, contemplando sua forma, seu conteúdo e o seu arranjo em acervos [...]" (SANTOS, P. 2008, p. 165-166), além das discussões sobre metadados definidos por Alves (2010), Alves e Santos (2013), Baca (2016), Pomerantz (2015), Zeng e Qin (2016), Joudrey, Taylor e Wisser (2018) e os estudos de proveniência dos dados tratada pelo *World Wide Web Consortium* (W3C). Assim, a partir dos conceitos apresentados, é possível estabelecer objetivos desta pesquisa.

1.5 Objetivos

Dessa forma, o **objetivo geral** é analisar a viabilidade da aplicação do modelo PROV-O para a representação da proveniência em registros bibliográficos de ambientes digitais.

Nesse contexto, os **objetivos específicos** configuram-se em:

- Realizar levantamento do conceito de proveniência em diversos contextos;
- Estabelecer as características dos metadados de proveniência e verificar a relação com as demais tipologias de metadados;
- Apresentar as características e as relações das especificações da família PROV;
- Identificar a proveniência nos padrões de metadados aplicados ao domínio bibliográfico (MARC21, *Dublin Core*, BIBFRAME, PREMIS e *Schema.org*).

1.6 Procedimentos metodológicos

Esta pesquisa caracteriza-se por uma pesquisa qualitativa, em razão da análise que busca entender o relacionamento entre os metadados administrativos, a proveniência dos

dados e a ligação de dados abertos no domínio bibliográfico. Com o intuito de elucidar o problema proposto, caracterizando por uma pesquisa exploratória conforme a concepção de Gil (2010).

A análise exploratória da literatura disponível sobre a proveniência permitiu a construção de um conhecimento teórico sobre os metadados administrativos. Essa análise foi realizada pelo estudo sobre a proveniência em diversos contextos, auxiliando na compreensão do problema proposto e na construção dos resultados esperados. Dessa forma, foi realizada uma pesquisa bibliográfica. Os temas abordados foram: Metadados administrativos e proveniência dos dados, tipologia de metadados, proveniência.

Nesse sentido, a busca pelo estado da arte como procedimento metodológico está em "[...] mapear e de discutir uma certa produção acadêmica em diferentes campos do conhecimento [...]" (FERREIRA, 2002, p. 258). Nesse contexto, o interesse está em identificar as pesquisas tanto do campo da Ciência da Informação quanto das áreas de intersecção, como Ciência da Computação, que também abordam preocupações sobre a proveniência.

O intuito do levantamento foca-se em "[...] responder que aspectos e dimensões vêm sendo destacados e privilegiados em diferentes épocas e lugares [...]" (FERREIRA, 2002, p. 258). Assim sendo, o recorte da pesquisa abrange pesquisas publicadas internacionalmente e no Brasil, nos idiomas em português, espanhol e inglês.

Os procedimentos para identificação do estado da arte segundo Nóbrega-Therrien e Therrien (2004, p. 08) se inserem "[...] em resumos e catálogos de fontes relacionados a um campo de investigação." Ferreira (2002) destaca ainda que são utilizados os resumos de dissertações de mestrado, teses de doutorado, artigos publicados em periódicos, relatórios e trabalhos apresentados nos principais eventos na área da Ciência da Informação e Computação.

A partir do estudo exploratório da literatura, foi possível concretizar os objetivos propostos sobre o universo da pesquisa e chegar a soluções e considerações acerca do problema de pesquisa. Deste modo, os procedimentos metodológicos foram divididos nas seguintes etapas:

1ª Etapa - Levantamento bibliográfico: o levantamento bibliográfico foi realizado em nível nacional e internacional e em fontes bibliográficas da área de estudo, a partir de bases de dados como específicas como:

- Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI);
- Repositório Questões em Rede – Coleções (Coleção BENANCIB);
- *Library and Information Science Abstracts* (LISA);
- Library, Information Science and Technology Abstracts (LISTA).

Bases de dados gerais:

- Biblioteca Digital Brasileira de Teses e Dissertações (BDTD);
- Teses e dissertações da CAPES;
- Oasisbr: portal brasileiro de publicações científicas em acesso aberto do Ibict;
- P@rthenon;
- Portal de Periódicos da Capes;
- *Scientific Electronic Library Online* (SciELO);
- *Scopus;*
- *Web of Science*.

Além de sites:

- Anais do *Dublin Core Metadata Initiative International Conference on Dublin Core and Metadata Applications*;
- *Google Scholar*.

Para realização do levantamento, foram elaboradas estratégias de buscas para localização dos materiais selecionados. Destaca-se que algumas bases de dados não foram possíveis utilizar a estratégia de busca exata, tendo que ser adaptada a partir da busca individualizada das palavras-chave. Quando possível, buscou identificar fazer a busca em título, palavras-chave e resumos.

- **Estratégia de busca para proveniência:** Provenance OR Proveniência OR Prov*
- **Estratégia de busca para questão das tipologias de metadados:** Metadatos OR Metadata OR Metadados
- **Estratégia de busca para trabalhos correlatos: (**Prov*) AND (Metadata OR MARC OR Dublin Core OR BIBFRAME OR Schema.org OR PREMIS)
- **Período levantado:** última década, ou seja, 2008-2018.

- **Tipologia**: Artigos, livros, teses, dissertações, trabalhos em eventos

2ª Etapa - Seleção do material obtido: Após o levantamento e a identificação do corpus do trabalho, foi realizado uma leitura prévia do resumo e quando necessário, uma leitura prévia do texto para que pudesse aplicar os seguintes critérios para seleção do material para fundamentação teórica do texto:

- a partir de uma leitura prévia do resumo, foi verificado a relevância da temática do artigo para o escopo da pesquisa;
- pertinência dos autores e periódicos para temática, preferencialmente para os periódicos qualificados pelo *Web*qualis, e indicadores do periódico da *Journal Citation Reports* (JCR) e *Scientific Journal Rankings* (SJR);
- idioma dos documentos (português, inglês e espanhol); e
- atualidade dos documentos.

Outros documentos foram necessários incluir na pesquisa, devido a pertinência para composição do referencial teórico da pesquisa, como também documentos e normas propostas de *sites* de instituições como *World Wide Web Consortium* (W3C) e *International Federation of Library Associations and Institutions* (IFLA).

3ª Etapa – Leitura, interpretação e fichamento: foi realizada leitura e fichamento dos documentos selecionados e organizando de forma lógica o conteúdo, reunindo documentos com temáticas similares. Essa etapa teve como propósito o desenvolvimento da base teórica para a discussão dos diferentes pontos de vista identificados na literatura sobre o tema.

4ª Etapa – Análise e estabelecimento das características fundamentais extraídas da literatura e sistematização do estudo exploratório: análise das principais características encontradas na literatura sobre a proveniência, metadados administrativos, proveniência dos dados para elucidação do problema de pesquisa, criando assim a base teórica para elaboração (redação) da pesquisa.

5ª Etapa - *Crosswalk* dos padrões para PROV: Após a compreensão do contexto da proveniência, foi realizada a correspondência dos principais padrões do domínio bibliográfico para o PROV. Segundo St. Pierre e LaPlant (1999), o processo de

correspondência de um padrão de metadados para outro é denominado de *Crosswalk*.

O método *crosswalk* é utilizado como processo para viabilizar a interoperabilidade entre sistemas que utilizam padrões de metadados heterogêneos e que precisam identificar a semântica estabelecida de cada metadado do padrão de metadado e mapear para o metadado que possui semântica igual ou similar. "*Crosswalks* fornece capacidade de fazer o conteúdo de elementos definidos em um padrão de metadados disponíveis para as comunidades que utilizam padrões de metadados relacionados." (ST.PIERRE; LAPLANT, 1999, tradução nossa).

Entretanto, destaca-se que, a correspondência por meio do *crosswalk* é difícil e propensa a erros, pois é necessário um conhecimento aprofundado dos padrões de metadados associados. (ST.PIERRE; LAPLANT, 1999, tradução nossa).

> *Crosswalks* não são importantes apenas para apoiar a exigência e facilidade que oferecem para discutir e organizar os serviços com os agentes da autoridade, ou pesquisas em domínios interdisciplinares, pois eles também são instrumentais para converter dados de um formato para outro que são mais acessíveis. (BACA, 2016, não paginado, tradução nossa).

St.Pierre e LaPlant (1999) sugerem etapas para realizar o *crosswalk*, que são divididos em 1) harmonização, 2) mapa semântico, 3) mapeando elemento a elemento, e 4) hierarquia, objeto e visão lógica, conforme apresentado no quadro 1.

Quadro 1 - Apresentação do método *Crosswalk*

Etapa	Subetapa	Observação
1° etapa: Harmonização Extração da terminologia comum, propriedades, organização e processos utilizados pelos padrões de metadados, e criar um quadro genérico para que se possa desenvolver novos ou rever padrões de	Subetapa A: Terminologia	Utilização de terminologias diferentes dos padrões dificultam o mapeamento entre eles.
		É essencial chegar a um acordo sobre a terminologia dos padrões, além de estabelecer uma definição formal para cada termo.
	Subetapa B: Propriedades - As semelhanças das propriedades dos padrões são extraídas e os conceitos generalizados.	Identificadores únicos para cada metadado, por exemplo, tag, etiqueta, identificador.
		Qual definição semântica de cada metadado?
		O metadado é obrigatório, opcional ou obrigatório em certas condições?
		Um metadado pode ocorrer várias vezes?
		Organização dos metadados em relação ao outro, por exemplo, as relações hierárquicas.
		Restrições impostas pelos valores do elemento (texto livre, escala numérica ou data)?
		Suporte opcional para elementos de metadados definidos localmente?
		As propriedades comuns podem ser expressas e utilizadas de uma

metadados já existentes.		forma similar dentro de cada padrão? Esta etapa simplifica o desenvolvimento do *Crosswalk*.
	Subetapa C: Organização	Para facilitar cada padrão deve ser organizado em forma similar, de modo que determinada seção de um padrão possa ser encontrada em uma seção de outro padrão.
	Subetapa D: Processo	Há ocasiões em que a escolha do processo selecionado seja arbitrária e não um processo análogo a outro padrão relacionado.
2° etapa: Mapa semântico	O mapeamento semântico é a especificação de cada elemento do padrão com o elemento semanticamente equivalente para o outro padrão. Para St.Pierre e LaPlant (1999) é o processo mais importante da harmonização e desenvolvimento do Crosswalk, pois determina o mapeamento semântico entre os padrões de metadados de origem e destino.	
3° etapa: Mapeando elemento a elemento - Identificar os metadados opcionais e obrigatórios. Nesta fase considerar as propriedades de cada metadado	**Uma para muitos:** ocorrência de vários elementos de origem a uma única ocorrência no elemento alvo. A um elemento que se está verificando irá ser correspondente a diversos elementos do outro padrão de metadados.	
	Muitos para um: muitos elementos de um padrão de metadados para apenas um metadado no padrão de destino. Devem-se aproximar todos os elementos do primeiro metadado e indicar a um único elemento do outro padrão. Se a resolução é mapear todos os valores do elemento de origem para um único valor no elemento alvo, regras explícitas são obrigadas a especificar como os valores serão anexados juntos. Caso seja apenas mapear um valor de elemento de origem para o destino, com a possível consequência de perda de informações, a resolução deve indicar os critérios para a seleção de elementos.	
	Elementos extras na fonte: Outro caso importante que requer resolução é a manipulação de um elemento de origem que não é mapeado para qualquer elemento apropriado no padrão alvo. Uma vez que muitos padrões fornecem a capacidade de capturar informações adicionais, a resolução deve especificar exatamente como o valor do elemento deve ser adicionado.	
	Elementos obrigatórios /não resolvidos em alvo: Em alguns casos, pode haver elementos obrigatórios no alvo que não têm mapeamento correspondente no padrão de metadados de origem. Porque o alvo requer um valor para os elementos obrigatórios, o *Crosswalk* deve fornecer uma resolução para os seus valores.	
4° etapa: Hierarquia, Objeto e Visão Lógica	**Hierarquia:** A maioria dos padrões de metadados organizam seus metadados hierarquicamente. Em alguns casos, a profundidade da hierarquia pode ser fixada. Em outros casos a profundidade da hierarquia é ilimitada.	
	Objeto: Item versus coleção. Item é um único documento, ou seja, os metadados associados a um documento. Coleção conjunto de itens, ou seja, os metadados referem-se a mais de um item.	
	Visão Lógica: Permite ver um conjunto específico de metadados do padrão organizado de uma maneira específica	
	Conversão de conteúdo: Padrões de metadados restringem o conteúdo de cada metadado para um determinado tipo de dado, intervalo de valores, ou vocabulário controlado. Muitas vezes, as conversões são baseadas não só nas propriedades que definem a fonte e os metadados alvo, mas também os conteúdos dos elementos de metadados de origem.	
	Combinações de conversão: Quando as propriedades de conversão são consideradas de forma independente, as conversões de metadados podem parecer simples para especificar e processar. Na prática, vários problemas de conversão refletem em uma combinação, o que dificulta a especificação de conversão e processo. Deve considerar as transformações necessárias para converter um metadado alvo, onde várias propriedades são diferentes do metadado de origem	

Fonte: Baseado em St.Pierre e LaPlant (1999)

Para realização do *crosswalk*, Chan e Zeng (2006) esclarecem que há duas possibilidades, o "*crosswalking* absoluto" e o "*crosswalking* relativo". O "*crosswalking* absoluto" é a correspondência exata entre os metadados, ou seja, a semântica de um

metadado é exatamente a mesma do metadado do outro padrão que está sendo analisado.

A correspondência garante equivalência dos elementos. Quando isso não ocorre no processo de correspondência dos metadados, não há o *crosswalking*, causando perdas de informações. Para minimizar esse problema, as autoras sugerem a realização do "*crosswalking* relativo", usado para corresponder os elementos de um esquema de fonte de pelo menos um elemento de um esquema de destino.

De acordo com Chan e Zeng (2006) o processo de *crosswalk* pode apresentar algumas dificuldades em diferentes graus de equivalência como: um-para-um, um-para-muitos, muitos-para-um e um-para-nenhum. A equivalência um-para-um corresponde à um metadado corresponde a apenas um metadado no outro padrão. Já na equivalência um-para-muitos significa que um metadado do primeiro padrão, pode ter diversos metadados com contexto similar, fazendo com que a correspondência final possa apresentar diversos metadados correspondidos. Na equivalência muitos-para-um significa que muitos metadados corresponde a apenas um metadado no segundo padrão. Por fim, a equivalência um-para-nenhum representa que um metadado não teve nenhum metadado correspondido com o segundo padrão.

Nesse contexto, optou-se em realizar o *crosswalk* do PROV-O para os padrões aplicados ao domínio bibliográfico, pois o intuito da pesquisa foi verificar a compatibilidade do PROV-O nos padrões do domínio bibliográfico e dessa forma, não apresenta a correspondência inversa, ou seja, dos padrões aplicados ao domínio bibliográfico para o PROV-O.

Para apresentação dos resultados, optou-se por realizar primeiro o *crosswalk* individual devido a complexidade dos padrões analisados, contendo uma coluna com as classes e propriedades do PROV-O, o padrão analisado e uma coluna com comentários.

Posteriormente, foi criado um quadro com o *crosswalk* do PROV-O para todos os padrões analisados.

6ª Etapa – Sistematização do conteúdo: A partir do levantamento bibliográfico, foram comparados os padrões de metadados e extraídos os metadados administrativos e mapeados com as especificações PROV, possibilitando assim, subsídios para ratificação da hipótese.

1.7 Estrutura da dissertação

Esta tese apresenta quanto à organização, além do presente capítulo que aborda questões iniciais e a contextualização da temática proposta: a importância dos metadados administrativos e da proveniência no reuso dados. Apresenta também a definição do problema, a motivação da execução da tese, os objetivos gerais e específicos e a justificativa apontando a relevância do trabalho.

CAPÍTULO 2 – CONTEXTO DA PROVENIÊNCIA. No capítulo foi apresentada uma revisão de literatura sobre Proveniência em diversos contextos.

CAPÍTULO 3 – FAMÍLIA PROV. Foi apresentado neste capítulo, um estudo dos documentos da família PROV e suas possíveis aplicações para descrições.

CAPÍTULO 4 – METADADOS E SUAS TIPOLOGIAS. Neste capítulo foi apresentado uma discussão sobre os metadados e suas tipologias.

CAPÍTULO 5 – PROVENIÊNCIA NOS PADRÕES DE METADADOS NO DOMÍNIO BIBLIOGRÁRICO. Após o referencial teórico, foi realizado o C*rosswalk* entre os principais padrões de metadados aplicados ao domínio bibliográfico, MARC21, *Dublin Core*, BIBFRAME, *Schema.org* e PREMIS.

CAPÍTULO 6 – DISCUSSÃO E ANÁLISE DOS DADOS. Neste capítulo foram apresentados os principais resultados e análise de dados.

CAPÍTULO 7 –CONSIDERAÇÕES FINAIS. Apresentação das considerações da tese.

REFERÊNCIAS – Descrição dos recursos utilizados durante a tese.

2 CONTEXTO DA PROVENIÊNCIA

O propósito deste capítulo é discutir o conceito proveniência em diversos domínios e suas características. Em geral, o termo proveniência tem sido empregado na identificação do responsável pela criação, guarda e gerenciamento de informações e recursos das mais diversas áreas. Segundo Moreau e Groth (2013, p. 4, tradução nossa), “A proveniência é definida como um registro que descreve as pessoas, instituições, entidades e atividades envolvidas na produção, influência ou entrega de um dado ou coisa.” De acordo com Pearce-Moses e Baty (2005, p. 317, tradução nossa) a proveniência é definida como “1. A origem ou fonte de alguma coisa. - 2. Informações sobre a custódia das origens e propriedade de um item ou coleção.”

Para Gil e Miles (2013) a proveniência pode representar diferentes perspectivas e tipos de informação. A primeira perspectiva está relacionada ao agente, ou criador, podendo ser uma pessoa ou organização, garantindo assim quem criou o recurso informacional. A segunda perspectiva está atrelada ao próprio recurso informacional, identificando por exemplo, sua origem. A terceira perspectiva está relacionada ao processo, registrando as ações e etapas tomadas para construção do registro informacional.

A proveniência é utilizada em diversas áreas como no jornalismo, nas artes, na museologia, na Arquivologia, na computação entre outras áreas para garantir a autenticidade das informações prestadas. Segundo Gil e Miles (2013)

> A proveniência pode ser usada para muitos propósitos, como entender como os dados foram coletados para que possam ser usados de forma significativa, determinar propriedade e direitos sobre um objeto, fazer julgamentos sobre informações para determinar se confiar nele, verificar se o processo e as etapas usadas para obter um resultado está em conformidade com determinados requisitos e reproduzindo como algo foi gerado.

Em especial, na Arquivologia, a proveniência tem o papel ainda de auxiliar na organização dos acervos. A proveniência na Arquivologia garante a organicidade do fundo arquivístico a partir do produtor do documento, tornando-se assim, parte de um conjunto de documentos criados por uma mesma instituição ou pessoa. De acordo com Pearce-Moses e Baty (2005, p. 317) a proveniência é um princípio fundamental dos arquivos, referindo-se ao indivíduo, família ou organização que criou ou recebeu os itens em uma coleção. O princípio da proveniência ou o respeito dos fundos corresponde que registros de diferentes origens

(procedência) sejam mantidos separados para preservar seu contexto. Segundo Duranti et al. (2016, p. 11, tradução nossa) “Os relacionamentos entre registros e as organizações ou indivíduos que criaram, acumularam e/ou mantiveram e usaram na condução de atividades pessoais ou corporativas.”

De acordo com o Arquivo Nacional (2015, p. 140), a proveniência é um “Termo que serve para indicar a entidade coletiva, entidade coletiva pessoa ou família produtora de arquivo.” A proveniência é fundamental para organização do fundo arquivístico, assim, o princípio da proveniência garante a integridade arquivística e consistência do fundo arquivístico. Segundo o Arquivo Nacional (2015, p. 136) o princípio da proveniência é “Princípio básico da arquivologia segundo o qual o arquivo produzido por uma entidade coletiva, pessoa ou família não deve ser misturado aos de outras entidades produtoras. Também chamado princípio do respeito aos fundos.” Ao analisar os trabalhos que falam sobre proveniência na arquivologia, Macedo (2018) realizou um levantamento da ocorrência do termo proveniência e observou escassez na utilização de termos e definições na arquivologia, evidenciando assim, a necessidade de estudos mais aprofundados sobre a temática.

Assim como na Arquivologia, em outras disciplinas, como a museologia, a proveniência é fundamental para garantir a procedência do item museológico ou pintura, para garantir sua autenticidade como objeto único.

> A proveniência é algo que dominou o comércio no mundo da arte. A ideia de que uma pintura é o que ela pretende ser (por exemplo, saber por quem ela foi pintada, quando, que não é uma falsificação ou cópia) afetará seu valor percebido. Essa idéia também se aplica a livros impressos e outros artefatos físicos, onde pode haver um valor associado a um manuscrito original ou a uma primeira edição. Essa ideia foi adotada no mundo comercial e se aplica à documentação associada a transações comerciais. (HAYNES, 2018, p. 134, tradução nossa)

Com a facilidade de copiar e reproduzir arquivos na ambiência digital, a proveniência é fundamental para dar autenticidade às informações contidas nos documentos. Segundo Haynes (2018, p. 134, tradução nossa) “Quando se trata de estabelecer a autenticidade de um item, sua história torna-se importante, sua proveniência: as circunstâncias de sua criação, quem a possuiu e as condições sob as quais sua propriedade foi transferida.”

Na percepção da preservação digital, a proveniência é vital para registrar informações dos responsáveis pela criação, custódia, alteração, curadoria e administração

do objeto digital. Haynes (2018, p. 134, tradução nossa) complementa que “No contexto dos materiais digitais, fornecer informações de proveniência pode ajudar a demonstrar que um registro não foi adulterado e que a evidência que ele apresenta é, portanto, confiável.” De acordo com o dicionário da *Library of Congress* conjunto ao *Premis Editorial Committee* (2018, p. 211, tradução nossa),

> Proveniência digital: documentação de processos no ciclo de vida de um objeto digital. Proveniência Digital tipicamente descreve Agentes responsáveis pela custódia e administração de Objetos Digitais, eventos-chave que ocorrem ao longo do ciclo de vida do objeto digital e outras informações associadas à criação, gerenciamento e preservação do objeto digital.

Paralelo às questões de preservação digital, no ambiente *Web*, a proveniência é fundamental para identificação e confiabilidade das informações compartilhadas. Segundo Moreau e Groth, 2013, p. 4, tradução nossa),

> No contexto da *Web*, proveniência é um registro que pode ser criado, trocado e processado por computadores. [...] O registro processável por computador contém descrições dos eventos ocorridos, levando para um recurso ou uma coisa, como existe em algum contexto. Muitos fatores podem contribuir para tal estado de assuntos, incluindo as pessoas envolvidas, as organizações em que atuam em nome dos processos que estão sendo executados e outros dados, recursos ou coisas que fazem parte dele.

Nesse contexto, Haynes (2018, p. 134, tradução nossa) argumenta que “O gerenciamento de registros e a boa governança dependem da capacidade de demonstrar a autenticidade de um registro, e fornecer a documentação sobre seu histórico e a maneira como ele foi gerenciado.” As informações necessárias para garantir a proveniência e autenticidades dos registros podem incluir quem e quando um determinado recurso informacional foi acessado, quais mudanças foram realizadas, entre outras informações.

O registro das atualizações de um recurso é fundamental para garantir a autenticidade das informações prestadas. O controle de versão, quem fez a atualização e quando são informações que devem persistir no registro informacional do recurso. Dessa forma, os metadados são fundamentais pois,

> [...] podem fornecer um registro da proveniência de um documento e evidências de que ele foi mantido para estabelecer padrões e seguir procedimentos definidos. Isso é vital para documentos que foram escaneados e digitalizados e onde o original foi destruído, bem como os documentos nascidos digitais. (HAYNES, 2018, p. 134-135, tradução nossa).

Segundo Hayes (2018), os metadados auxiliam na veracidade da integridade dos recursos informacionais digitais, no mesmo contexto que eram estabelecidas as autenticidades de recursos informacionais em papel como contratos e testamentos.

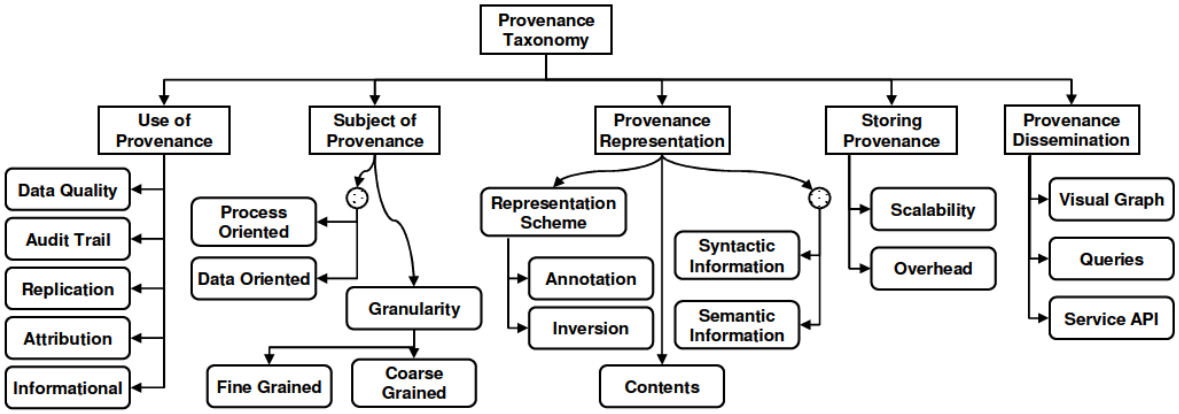
> Tradicionalmente, a autenticidade de documentos com peso legal, assim como contratos e testamentos, era estabelecida pela assinatura e uma marca de identificação, como um selo ou marca d'água. Eles também tinham metadados associados a eles, como detalhes de como o documento havia sido mantido e informações sobre os procedimentos para prevenir adulterações ou mesmo mudanças de qualquer rei para o original. (HAYNES, 2018, p. 134-135, tradução nossa).

No contexto da Ciência da Computação, a proveniência de dados é tratada por Buneman, Khanna e Tan (2000, p. 316, tradução nossa) como "[...] a descrição das origens de um dado e o processo pelo qual ele chegou em um banco de dados." A proveniência tem sido abordada principalmente no desenvolvimento de fluxos de trabalhos (*workflows*) para garantia da proveniência de dados científicos. Conforme apontado pelos autores, o reuso de dados sem a devida identificação pode resultar um problema sério, como a não identificação da proveniência das informações prestadas.

> O campo da biologia molecular, por exemplo, suporta cerca de 500 bancos de dados públicos, mas apenas alguns deles são dados "de origem", no sentido de que recebem dados experimentais. Todos os outros bancos de dados são, em algum sentido, visualizações dos dados de origem ou de outras visualizações. De fato, alguns deles são visões um do outro, o que parece sem sentido até que se entenda que os bancos de dados individuais não são simplesmente computados por consultas, mas também têm valor agregado na forma de correções e anotações por especialistas (eles são "curados"). Um problema sério enfrentado pelo usuário de um desses bancos de dados é conhecer a proveniência de um determinado dado. Esta informação é essencial para qualquer pessoa interessada na precisão e pontualidade dos dados. (BUNEMAN; KHANNA; TAN, 2000, p. 316, tradução nossa).

Pensando nos requisitos de proveniência, Simmhan, Plale e Gannon (2005) estabeleceram uma taxonomia das diversas possibilidades de informações da proveniência, sendo que *use of provenance* (uso da proveniência), *subject of provenance* (objeto da proveniência), *provenance representation* (representação da proveniência), *storing provenance* (armazenamento da proveniência) e *provenance dissemination* (disseminação da proveniência) representam categorias conforme apresentado na figura 1.

Figura 1 - Taxonomia da proveniência



Fonte: Simmhan, Plale, Gannon (2005, p. 33)

Na categoria "*Use of Provenance*" (uso da proveniência), de acordo com Simmhan, Plale e Gannon (2005, p. 33), está relacionada à *Data quality* (qualidade de dados), *Audit trail* (trilha da auditoria), *Replication* (replicação), *Attribuition* (atribuição) e *Informational* (informacional). A Qualidade de dados pode ser usada para estimar a qualidade e confiabilidade dos dados com base na origem e transformações dos dados. Também pode fornecer declarações sobre derivação dos dados. Na subcategoria Trilha de Auditoria, a proveniência pode ser usada para rastrear a trilha dos dados, determinar o uso de recursos e detectar erros na geração de dados. Replicação está relacionada às informações de proveniência que permitem a repetição da derivação de dados. A atribuição estabelece os direitos autorais e propriedade de dados, permite sua citação e determina a responsabilidade em caso de dados errados. Já na subcategoria Informacional, está relacionada à consulta nos metadados para descoberta de dados da linhagem do recurso.

Na categoria "*Subject of Provenance*" há informações de proveniência coletadas de diferentes recursos no sistema de processamento orientado aos dados, em processos e em níveis de detalhe. No modelo orientado aos dados, os metadados são reunidos especificamente sobre os dados. Já no modelo orientado aos processos, os processos derivados são as entidades primárias cuja proveniência é coletada e a proveniência dos dados é determinada a partir da inspeção dos dados de entrada e saída destes processos.

O nível de detalhes é conhecido também como granularidade. Segundo os trabalhos de Woodley, Clement e Winn (2005) e Alves, Simionato e Santos (2012), o termo granularidade está relacionado ao nível de detalhe em que um objeto ou recurso é visto ou

descrito. Santos e Sant'Ana (2013, p. 206) complementam que "[...] a granularidade de um conjunto de dados está vinculada ao número de atributos que o compõem e a diversidade de seus conteúdos."

Nesse sentido, a granularidade está vinculada ao conteúdo disponível no conjunto de dados e impacta diretamente nos processos de acesso e de tratamento da informação. (SANTOS; SANT'ANA, 2013, p. 207). Basicamente a granularidade é dividida em dois níveis: "[…] a granularidade fina (*fine granularity*) significa que a descrição apresenta um alto nível de detalhamento; e a granularidade grossa (*coarse granularity*) significa que a descrição possui um baixo nível de detalhamento." (SIMIONATO, 2015, p. 72-73).

> A utilidade da proveniência em um determinado domínio é ligada à granularidade em que é coletado. Os requisitos variam de proveniência em atributos e tuplas numa base de dados para prover coleções de arquivos, digamos, gerados por um experimento conjunto. (SIMMHAN; PLALE; GANNON, 2005, p. 33, tradução nossa).

De acordo com Simmhan, Plale e Gannon (2005, p. 33) a "*provenance representation*" (representação de proveniência) pode ser registrada a partir de diversas informações que acarretam em implicações no custo de registrá-la e na riqueza de seu uso. Há duas formas de representar a proveniência, por "*annotation*" (anotação) ou "*inversion*" (inversão). Nas Anotações, são coletados o histórico da derivação dos dados. Já o método de inversão, "[...] usa a propriedade pela qual algumas derivações podem ser invertidas para encontrar os dados de entrada fornecidos a eles para derivar os dados de saída." (SIMMHAN; PLALE; GANNON, 2005, p. 33, tradução nossa).

Apesar dos autores Simmhan, Plale e Gannon (2005) não deixarem claro, há outras subcategorias da representação estão: *syntactic information* (informação sintática), *semantic information* (informação semântica) e *contents* (conteúdo). A informação sintática está relacionada à estrutura das informações descritas como linguagens de marcação, por exemplo, *Extensible Markup Language* (XML), *Turtle, Jason*, entre outras. As informações semânticas estão atreladas à representação por meio de ontologias, e o conteúdo está contido nas relações entre as informações capturadas.

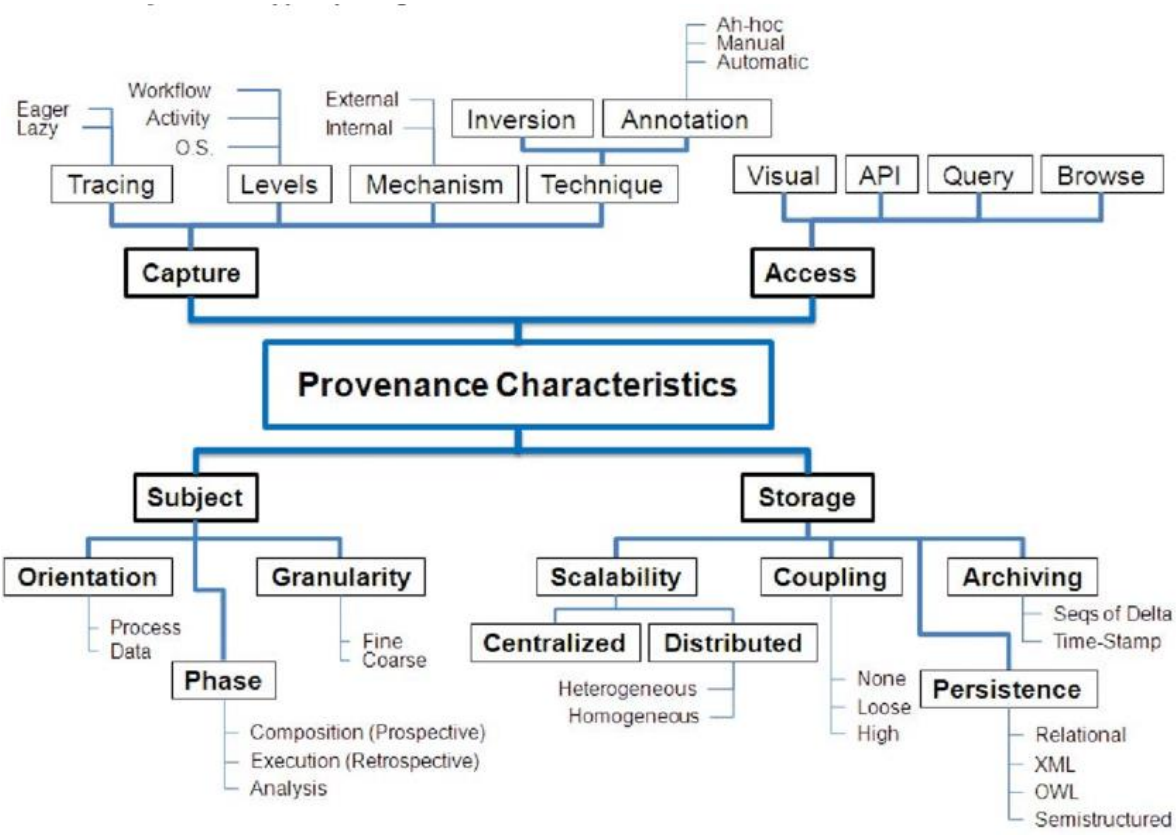
O "*Storing Information*" (Armazenamento de Proveniência) corresponde à forma pela qual os metadados são armazenados e podem ser maiores que os dados descritivos se os dados tiverem uma granularidade fina e as informações de proveniência forem ricas e exaustivas. A representação impacta diretamente no armazenamento, pois utilizar o método

de inversão que possui um conjunto de informações com uma granularidade grossa é mais fácil armazenar, enquanto o método de anotações é mais detalhado. (SIMMHAN; PLALE; GANNON, 2005, p. 33). *Scalability* (Escalabilidade) e *Overhead* (sobrecarga) são subcategorias do armazenamento. Escalabilidade pode ser definida como capacidade de gerenciar uma quantidade crescente de informações. A sobrecarga está relacionada à capacidade de armazenamento das informações.

Outra subcategoria definida por Simmhan, Plale e Gannon (2005) é a "*Dissemination Information*" (Divulgação da Proveniência), que corresponde na possibilidade dos sistemas permitirem formas de acessar a proveniência. Como subcategoria, pode ser por *Visual Graph* (Visualização de gráficos) que possibilita o usuário navegar e inspecionar elementos, *Queries* (consultas) que os usuários possam fazer buscas em conjuntos de dados com base nos metadados de proveniência, e ainda, *Service API* (Serviços API), que permitem os usuários a criarem e implementar seus próprios mecanismo de recuperação via *Application Programming Interface* (API), ou seja, programas de aplicação em interfaces.

Os autores Cruz, Campos e Mattoso (2009), fizeram um estudo das características da proveniência no contexto dos *Scientific Workflow Management Systems* (SWfMS), conforme apresentado na figura 2.

Figura 2 – Características do Sistema de Taxonomia de Proveniência



Fonte: Cruz, Campos e Mattoso (2009, p. 263).

Diferente do trabalho de Simmhan, Plale e Gannon (2005) que possui uma concepção mais genérica, Cruz, Campos e Mattoso (2009, p. 263) buscaram especificar a taxonomia para ciclo de vida do fluxo de trabalhos científicos, abordando ainda as concepções do *Open Provenance Model* (OPM)[1]. Nesse contexto, a taxonomia apresentada por Cruz, Campos e Mattoso (2009) apresenta quatro categorias maiores *capture* (captura)*, access* (acesso)*, subject* (assunto) e *storage* (armazenamento).

De acordo com Cruz, Campos e Mattoso (2009), na categoria *capture* estão relacionadas as seguintes subcategorias: *Tracing* (Rastreamento) ou origem do rastreamento pode ser feito por duas formas: *Eager* e *Lazy*. Na primeira (*Eager*) calcula a proveniência de um dado produto somente quando necessário, caso a caso, já no segundo caso (*Lazy*), ele "[...] calcula a proveniência dos dados imediatamente carregando a proveniência do produto de dados juntamente com a transformação dos dados, tornando-o pronto para ser usado como metadados." (CRUZ; CAMPOS; MATTOSO, 2009, p. 261, tradução nossa).

1 OPM é um modelo de dados interoperável para proveniência e será abordado no capítulo 4.

Outra subcategoria é *Levels*, ou níveis de captura, que pode ser dividido em três formas: *Workflow* (fluxo de trabalho)*, activity* (níveis de atividades), O.S. (Sistema operacional). A proveniência no nível de *Workflow* é a mais comum das abordagens, a partir de um mecanismo de obtenção e coleta e armazenamento de todos os dados de proveniência que são anexados ou integrados em um SWfMS. (CRUZ; CAMPOS; MATTOSO, 2009). Os níveis de atividade "[...] requer cada serviço ou processo envolvido em uma tarefa computacional para capturar sua própria informação de proveniência." (CRUZ; CAMPOS; MATTOSO, 2009, p. 261, tradução nossa). Entre as formas de captura é a que possui menor granularidade, da coleta. No caso da captura por sistema operacional, a proveniência é coletada no nível do sistema da *Application Programming Interface* (API)[2], eles dependem da disponibilidade de funcionalidade específica no nível do sistema operacional. (CRUZ; CAMPOS; MATTOSO, 2009).

De acordo com Cruz, Campos e Mattoso (2009) existem dois mecanismos (*Mechanism*) principais para capturar a proveniência, *External* (externa) ou *Internal* (interna). "Alguns usam estruturas internas para capturar a procedência informação, alguns dependem de serviços externos, que podem ser bastante genéricos, para recolher proveniência de distribuição e ambientes heterogêneos." (CRUZ; CAMPOS; MATTOSO, 2009, p. 261).

Na subcategoria *Technique* (Técnicas de captura de proveniência), refere-se a técnicas para capturar proveniência nos sistemas de proveniência e pode ser feita por *Inversion* (inversão) ou *Annotation* (anotações). Na inversão, os dados de proveniência podem ser usados para recriar resultados de produtos de dados que não podem ser acessados ou são muito caros para acessar. Segundo Cruz, Campos e Mattoso (2009, p. 262, tradução nossa) "Se uma gestão de sistema de proveniência é capaz de calcular a inversão de uma transformação (método de inversão), então a inversão pode ser usada para recriar itens de dados de origem dos itens de dados de resultado." A anotação de um produto de dados é a processo de adicionar ou "marcar" os dados existentes, ou seja, são metadados que descrevem procedimentos ou dados. As anotações podem ser fornecidas manualmente por usuários, automaticamente por aplicativos, ou inseridas via *ah-hoc*, e registram decisões e notas importantes em diferentes níveis de granularidade possibilitando assim, a compreensão do significado dos produtos de dados ou aplicações científicas. (CRUZ;

2 API em português Interface de Programação de Aplicações.

CAMPOS; MATTOSO, 2009).

A categoria *Access* possibilita o acesso por meio de um gráfico de derivação que permite os usuários procurarem os dados. Outra maneira viável é a consulta de dados por meio do uso de linguagens de consulta geral, podendo ser acessado ainda por API ou Navegação. (CRUZ; CAMPOS; MATTOSO, 2009).

Sujeito (*Subject*) da Proveniência, os dados de proveniência podem ser representados em termos de assuntos e níveis distintos de detalhes. A questão da granularidade corresponde na mesma percepção de Simmhan, Plale, Gannon (2005). Quanto a orientação (*Orientation*) corresponde à descrição do processo (ou sequências de etapas) que, juntamente com dados e parâmetros de entrada, levaram a criação do produto de dados, podem estar relacionadas ao processo ou aos dados. (CRUZ; CAMPOS; MATTOSO, 2009). As fases (*Phase*) de proveniência foram classificadas em três momentos distintos para captura de proveniência: *composition* (prospective), *execution* (retrospective), *analysis* (análise). Na fase de composição corresponde às etapas que precisam ser seguidas para gerar um produto de dados ou classe de produtos de dados. Na fase de execução pode-se coletar a "[...] procedência retrospectiva, como as gravações de quando e onde cada procedimento foi executado e como cada invocação se comportou; capta os passos que foram executados, bem como informações sobre a execução ambiente utilizado para derivar um produto de dados específico." (CRUZ; CAMPOS; MATTOSO, 2009, p. 262, tradução nossa). Já na fase de análise pode-se "[...] avaliar resultados científicos e as mudanças que foram feitas em uma especificação de fluxo de trabalho, consultando dados de proveniência como diferentes parâmetros usados para decretá-lo." (CRUZ; CAMPOS; MATTOSO, 2009, p. 262, tradução nossa). Nessa fase, os usuários podem ter acesso às outras duas fases anteriores.

Outro ponto apresentado por Cruz, Campos e Mattoso (2009) é o armazenamento (*Storage*). Segundo os autores, há diversos pontos a serem considerados no armazenamento. Um deles é a escalabilidade de armazenamento (*Scalability*) que está relacionada em como os dados irão estar armazenados, se será em um lugar centralizado (*Centralized*) ou distribuído (*Distributed*), ou seja, em diversos repositórios. Esses repositórios podem utilizar sistemas homogêneos (*Homogeneous*) ou heterogêneos (*Heterogeneous*).

Para garantir o armazenamento (*Storage*), há possibilidade de criar estratégias de acoplamento de três formas: *nocoupling* (sem acoplamento), *tight-coupling* (acoplado) e

loose-coupling (solto) acoplamento. A estratégia de *nocoupling* armazena informações de proveniência em um ou vários repositórios específicos de proveniência. A segunda opção, armazena a proveniência diretamente associada aos dados para os quais a proveniência é registrada, na figura 2, é denominado com o termo *High*. Já o acoplamento solto, usa um esquema de armazenamento misto, onde os itens de dados de proveniência e experimento são armazenados em um sistema de armazenamento, mas separados logicamente. (CRUZ; CAMPOS; MATTOSO, 2009).

As estratégias de armazenamento da proveniência poder ser por Sequência-delta (*seqs of delta*) e registro de alterações de tempo (*time-stamp*). A sequência de delta armazena uma versão dos itens de dados (ou seja, a primeira versão ou última versão) ou uma sequência para frente ou para trás entre versões sucessivas. A sequência de delta e o registro de data e hora divergem apenas em sua mudança representação: a técnica de sequência de delta descreve a diferença entre duas versões ao longo do tempo, já a segunda técnica descreve a diferença em torno dos dados de alteração da hora. (CRUZ; CAMPOS; MATTOSO, 2009).

Outro aspecto importante é a questão da persistência da proveniência (*Persistence*). Esta categoria descreve a expressividade dos modelos conceituais utilizados por um sistema de gestão de proveniência para persistir a proveniência. (CRUZ; CAMPOS; MATTOSO, 2009).

Conforme observado no contexto científico, os metadados de proveniência são importantes para reprodutibilidade dos procedimentos de uma pesquisa. Com a propagação da *e-science* e a disponibilização de conjunto de dados abertos em repositórios digitais é de vital importância uma representação adequada, principalmente para a equipe envolvida no levantamento dos dados, que precisa saber quando foi realizado levantamento e se houve alteração após sua disponibilização na internet.

> Enquanto a ciência está se tornando intensiva em computação e dados, o princípio fundamental do método científico permanece inalterado: experimental os resultados precisam ser reproduzíveis. Em contraste com um fluxo de trabalho, que pode ser visto como uma receita que pode ser aplicado no futuro, a proveniência é considerada equivalente a um diário de bordo, capturando todas as etapas envolvidas na derivação real de um resultado e que poderiam ser usadas para repetir a execução que levou a esse resultado, a fim de validá-lo. [...] Como Weitzner observa, a proveniência é um substrato que pode ser usado para executar verificações de políticas e tornar os sistemas responsáveis. (MOREAU; GROTH, 2013, p. 3, tradução nossa).

No contexto dos repositórios digitais, alguns estudos como de Vidotti et al. (2017) destacaram a importância de identificar a origem dos registros coletados pelo Repositório Institucional UNESP das diversas bases de dados, para identificar as principais fontes de cada área, com o intuito de ampliar o número de fontes para a coleta de registros.

Diante a importância da proveniência de dados no âmbito científico, foram criados workshops e conferências específicos como o *International Provenance and Annotation Workshop* (IPAW) e o *Provenance Challenge* para discutir a questão da proveniência. No evento *Provenance Challenge*, a comunidade científica tem a oportunidade de elencar e discutir os desafios de proveniência de dados a serem resolvidos. Como resultado do primeiro encontro em 2006, surgiu o modelo de proveniência digital, *Open Provenance Model* (OPM). (BIVAR et al., 2013, p. 2, tradução nossa).

A questão de proveniência também vem sendo destacada como tópicos de melhores práticas do *World Wide Web Consortium* (W3C). No início do ano de 2017, o W3C publicou uma recomendação sobre boas práticas para publicação de dados na *Web* "*Data on the Web Best Practices* W3C". O documento apresenta 14 tópicos, entre eles um específico que trata da importância da proveniência dos dados, tópico "8.4 *Data Provenance*" e categorizam quais os benefícios dessas práticas para publicação dos dados.

A proveniência é tratada na boa prática 5 "*Provide data provenance information*" e sugere-se que o fornecimento de informações completas sobre as origens dos dados e quaisquer alterações feitas. (LÓSCIO; BURLE; CALEGARI, 2017). "A proveniência é um meio pelo qual os consumidores de um conjunto de dados julgam sua qualidade. Entender sua origem e histórico ajuda a determinar se deve confiar nos dados e fornece um contexto interpretativo importante." (LÓSCIO; BURLE; CALEGARI, 2017, não paginado, tradução nossa).

De acordo com as melhores práticas de publicação de dados, para testar essa boa prática, é necessário verificar se "[...] os metadados do próprio conjunto de dados incluem as informações de proveniência sobre o conjunto de dados em um formato legível por humanos. Verifica-se um aplicativo de computador pode processar automaticamente as informações de proveniência sobre o conjunto de dados." (LÓSCIO; BURLE; CALEGARI, 2017, não paginado, tradução nossa). Os benefícios dessa boa prática correspondem à: Reuso, compreensão e confiança.

Para exemplificar a descrição da proveniência dos dados é apresentado no quadro 2,

como apresentar essas informações. No exemplo, “As propriedades dct:creator, dct:publishere dct:issued são utilizados para dar informações sobre a origem do conjunto de dados. A propriedade *prov:actedOnBehalfOf* é usada para designar que Adrian agiu em nome da Agência de Transporte da MyCity.” (LÓSCIO; BURLE; CALEGARI, 2017, não paginado, tradução nossa).

Quadro 2 - Exemplo de informações de proveniência

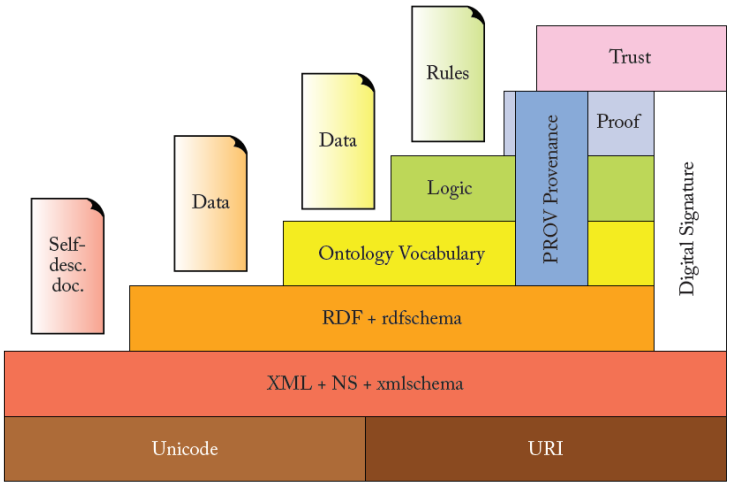
```
:stops-2015-05-05
    a dcat:Dataset, prov:Entity ;
    dct:title "Bus stops of MyCity" ;
    dcat:keyword "transport", "mobility", "bus" ;
    dct:issued "2015-05-05"^^xsd:date ;
    dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
    dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
    dct:spatial <http://sws.geonames.org/3399415> ;
    dct:publisher :transport-agency-mycity ;
      dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
      dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
      dct:creator :adrian

:adrian
    a foaf:Person, prov:Agent ;
    foaf:givenName "Adrian" ;
    foaf:mbox <mailto:adrian@mycitytransport.org> ;
    prov:actedOnBehalfOf :transport-agency-mycity
    .
:transport-agency-mycity
    a foaf:Organization, prov:Agent ;
    foaf:name "Transport Agency of Mycity"
```

Fonte: *World Wide Web Consortium* (2017, não paginado)

A proveniência tem ganhado destaque em discussões referentes à *Web* Semântica. A falta de registro das possibilidades de ligações e de relacionamento de dados, pode causar grandes problemas para o reuso de dados.

Moreau e Groth (2013) discutem a possibilidade de inclusão das questões de proveniência no bolo de noiva, proposto por Berners-Lee, Hendler, Lassila (2001), conforme apresentado na figura 3.

Figura 3 - Proveniência no diagrama de bolo de noiva da Web Semântica



Fonte: Moreau e Groth (2013, p. 4)

A camada da *Web* Semântica inicia-se com os Unicode e *Uniform Resource Identifier* (URI). O Unicode é a decodificação das informações contidas no registro e o URI representa identificações únicos para auxiliar na descrição e identificação das informações prestadas.

A segunda camada de baixo para cima, está relacionada aos formatos que compõem a estrutura da representação das informações, como *Extensible Markup Language* (XML), podendo ser utilizados ainda outros formatos como *Turtle*, e *Resource Description Framework in Attributes* (RDFa). Na terceira camada está no uso do *Resource Description Framework* (RDF) para auxiliar na troca de informações entre sistemas e seus relacionamentos. Na quarta camada aborda a questão das ontologias, que auxiliam na representação dos relacionamentos entre uma informação, e a partir dessas informações, realizar inferências para apresentar resultados contextualizados. A camada de lógica estabelece regras para expressar os relacionamentos propostos na ontologia. Já na camada de prova, está relacionada a veracidade ou não das informações prestadas. Paralelo a camada de interoperabilidade (RDF), lógica e prova está a assinatura digital, que busca mecanismo para validar as informações prestadas. Por fim, todas essas informações estabelecerão a confiança das informações prestadas. Entretanto, para os autores Moreau e Groth (2013), a proveniência perpassa as camadas de ontologia e lógica e está atrelada à camada *Proof,* ressaltando assim a importância da proveniência para dar confiabilidade às informações prestadas.

3 FAMÍLIA PROV

Neste capítulo aborda um breve histórico da concepção do *Provenance Working Group* do W3C, para discutir elementos para descrição da proveniência em recursos informacionais em âmbito geral. Fruto desses estudos foram criadas diversas recomendações e notas para melhor representar e descrever a proveniência. Assim, foram publicados diversos documentos que são conhecidos como documentos da família PROV. A família de documentos PROV conta com quatro recomendações, *The PROV Data Model* (PROV-DM), *The PROV Ontology* (PROV-O), *The Provenance Notation* (PROV-N) e *Constraints of the PROV Data Model* (PROV-CONSTRAINTS). Ainda foram publicadas oito (8) notas que auxiliam no mapeamento e nas informações sobre o modelo PROV.

3.1 *Background*: iniciativas para representação da proveniência

Nos últimos anos, algumas iniciativas têm explorado a questão de como representar a proveniência. O exemplo do *Open Provenance Model* (OPM), *Provenance, Authoring and Versioning* (PAV) e o PROV.

O OPM é um modelo conceitual de proveniência que define quais informações são necessárias em um sistema de proveniência. (CRUZ; CAMPOS; MATTOSO, 2009). As discussões da construção do OPM tiveram início em 2006, no primeiro *International Provenance and Annotation Workshop* (IPAW), mas só foi lançado para comunidade em 2007. A proposta do OPM foi definir um modelo de dados que seja aberto do ponto de vista da interoperabilidade, mas também com relação à comunidade de seus colaboradores, revisores e usuários. (MOREAU et al., 2011).

Entretanto, com o grupo de trabalho de proveniência da W3C com pesquisadores que estudavam o OPM, surgiu o PROV. De acordo com Moreau e Groth (2013) e Bivar et al. (2013, p. 2, tradução nossa) “O uso de proveniência, independente do modelo utilizado, fornece um fundamento essencial para avaliar a autenticidade de dados, permitindo confiabilidade e reprodutibilidade.” Para Bivar et al. (2013) com o surgimento do PROV, há uma probabilidade de migração dos sistemas que utilizam OPM para PROV, pois o PROV é apoiado por uma instituição de peso como o W3C. Nesse contexto, foi abordado o PROV

para discussões e construção do modelo desta tese.

De acordo com Haynes (2018, p. 135) “O PROV é um padrão para metadados de proveniência, que é hospitaleiro para fornecer metadados de outros esquemas. É baseado em um modelo de Agente, Entidade e Atividade [...]” e apresenta um modelo geral para representar informações de proveniência. Destaca-se que a proposta do padrão PROV não é abranger todas as especificidades de vários domínios, mas fornecer um conjunto de metadados para garantir um mínimo de informações de proveniência aplicável a todos domínios.

Já o PAV é um padrão focado na questão garantia da proveniência e identificação de pessoas e organizações e suas funções. Isto é, focado em quem criou, contribuiu e faz a curadoria dos dados e não faz uma abordagem específica dos processos conforme destacado por Ciccarese et al. (2013). Entretanto, não é o foco deste trabalho trabalhar especificamente com essas questões e o PROV apresenta um modelo mais geral sobre o Agente, Entidade e Atividade.

3.2 Visão geral da Família PROV

A família PROV foi criada a partir de um grupo de trabalho da World Wide Web (W3C) para discutir questões relativas à proveniência dos dados. "A família de documentos PROV define um modelo, serializações e outras definições de apoio correspondente que permitem o intercâmbio de informações de proveniência em ambientes heterogêneos como a *Web*." (GROTH; MOREAU, 2013, não paginado, tradução nossa). A família de documentos PROV conta com quatro recomendações, *The PROV Data Model* (PROV-DM), *The PROV Ontology* (PROV-O), *The Provenance Notation* (PROV-N) e *Constraints of the PROV Data Model* (PROV-CONSTRAINTS).

A recomendação PROV-O expressa o PROV-DM usando a *Web Ontology Language* (OWL) 2, fornecendo um conjunto de classes, propriedades e restrições que podem ser usadas para representar e trocar informações de procedência geradas em sistemas diferentes e em diferentes contextos. (LEBO; SAHOO; MCGUINNESS, 2013).

O PROV-N fornece uma notação de sintaxe compreensível tanto por máquina, quanto por seres humanos, com o intuito de descrever as instâncias do modelo de dados

PROV. (MOREAU; MISSIER, 2013a).

O PROV-CONSTRAINTS define um subconjunto de instâncias PROV chamado de instâncias PROV válidas, ajudando nas definições, inferências e restrições para outros padrões da *Web*. (NIES, 2013)

Ainda foram publicadas pelo W3C, oito (8) notas que auxiliam no mapeamento e informações sobre o modelo PROV. O PROV foi projetado para que usuários e desenvolvedores possam começar com o uso básico das questões de proveniência e, posteriormente, progredir progressivamente para cenários de uso mais avançados.

Para facilitar o entendimento do PROV, os documentos foram classificados com recomendações para três perfis de público. O primeiro grupo é destinado ao público geral de usuários, que queiram entender o PROV e usar aplicativos que suportam o PROV. O segundo grupo é para desenvolvedores que queiram criar ou desenvolver aplicativos que criem e consumam proveniência. Por fim, o terceiro grupo é denominado avançado que queira criar validadores, novas serializações ou outros sistemas baseados no PROV. (GROTH; MOREAU, 2013).

A relação dos documentos e público-alvo foram representados no quadro 3. Na coluna público, faz uma menção qual o público-alvo do documento. Na coluna Tipo, informa se o documento é uma nota ou uma recomendação (Rec). Já na coluna Documento, apresenta o documento com uma breve definição e função do documento.

Quadro 3 - Panorama das especificações da família PROV

Parte	Público	Tipo	Documento
1	Usuários	Nota	PROV-PRIMER é o ponto de início para o PROV, oferece uma introdução ao modelo de dados de proveniência. É aqui que você deve começar e, para muitos, pode ser o único documento necessário.
2	Desenvolvedores	Rec	O PROV-O define uma ontologia OWL2 leve para o modelo de dados de proveniência. Isso é destinado à comunidade Linked Data e Semantic *Web*.
3	Desenvolvedores	Nota	O PROV-XML define um esquema XML para o modelo de dados de proveniência. Isso é destinado a desenvolvedores que precisam de uma serialização XML nativa do modelo de dados PROV.
4	Avançado	Rec	O PROV-DM define um modelo de dados conceitual para a proveniência, incluindo diagramas UML. PROV-O, PROV-XML e PROV-N são serializações deste modelo conceitual.
5	Avançado	Rec	O PROV-N define uma notação legível para o modelo de proveniência. Isso é usado para fornecer exemplos dentro do modelo conceitual, bem como usado na definição de PROV-CONSTRAINTS.
6	Avançado	Rec	PROV-CONSTRAINTS define um conjunto de restrições no modelo de dados PROV que especifica uma noção de proveniência válida. Ele é especificamente destinado aos implementadores de validadores.

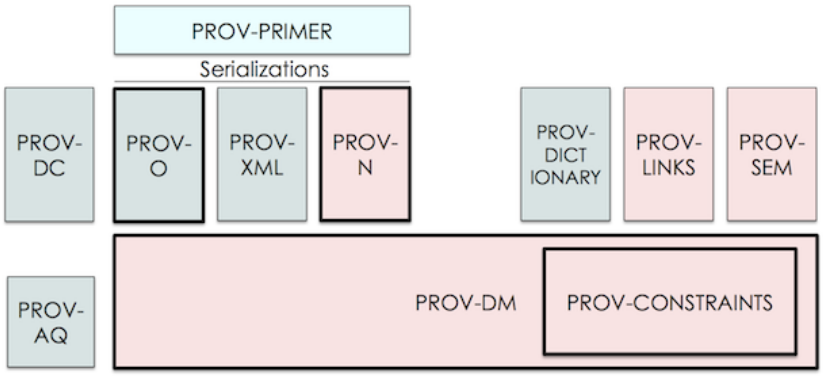
7	Desenvolvedores	Nota	O PROV-AQ define como usar mecanismos baseados na *Web* para localizar e recuperar informações de proveniência.
8	Desenvolvedores	Nota	O PROV-DC define um mapeamento entre o Dublin Core e o PROV-O.
9	Desenvolvedores	Nota	PROV-DICIONARY define construções para expressar a proveniência de estruturas de dados no estilo de dicionário.
10	Avançado	Nota	O PROV-SEM define uma especificação declarativa em termos de lógica de primeira ordem do modelo de dados PROV.
11	Avançado	Nota	O PROV-LINKS define extensões para o PROV para permitir a vinculação de informações de proveniência em pacotes de descrições de proveniência.

Fonte: Groth e Moreau (2013, não paginado, tradução nossa)

O documento PROV-PRIMER é destinado ao público em geral, denominado de "Usuários". Para desenvolvedores recomenda-se o aprofundamento dos documentos PROV-O, PROV-XML, PROV-AQ, PROV-DC e o PROV-*Dicionary*. Para o grupo avançado, recomenda o aprofundamento dos documentos: PROV-DM, PROV-N, PROV-*Constraints*, PROV-SEM e PROV-LINKS.

Para melhor visualizar a relação dos documentos destacados no quadro 3, a figura 4 apresenta os documentos da família PROV. O PROV-DM apresenta um modelo e é a base de todos os documentos. Os documentos em vermelho correspondem ao público avançado, o documento em azul (PROV-PRIMER) é para o público geral e os documentos em verde, são destinados aos desenvolvedores. O destaque do contorno do retângulo do documento representa que o documento é uma recomendação W3C.

Figura 4 - Família PROV



Fonte: Groth e Moreau (2013, não paginado)

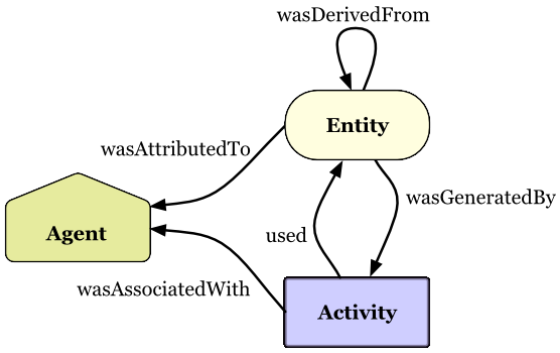
Nos próximos subcapítulos, serão apresentados os principais documentos da família PROV. Para o desenvolvimento da tese foram utilizados os documentos: PROV-OVERVIEW, PROV-PRIMER, PROV-O, PROV-DM e o PROV-DC. O PROV-OVERVIEW foi utilizado para

realização da introdução do PROV, o PROV-PRIMER foi utilizado para apresentar os principais conceitos relacionados à proveniência. Posteriormente, foi apresentado o modelo de dados proposto pelo PROV, o PROV-DM. Por conseguinte, foi trabalhado a ontologia PROV, com o PROV-O e ampliando as discussões do *Dublin Core* com o PROV, a partir do PROV-DC.

3.3 PROV-PRIMER

O PROV-PRIMER apresenta conceitos fundamentais para entender os demais documentos da família PROV, assim, ele serve como uma introdução e guia para o Modelo de dados PROV e para promover o intercâmbio da proveniência na *Web*.(GIL, Y; MILES, 2013). De acordo com Yolanda Gil e Miles (2013), o documento PROV-PRIMER é tido como documento base para entender o PROV e inicia-se apresentando os conceitos básicos do PROV. O diagrama apresentado na figura 5, fornece uma visão geral da estrutura dos registros PROV, limitada a alguns conceitos-chave de PROV.

Figura 5 - Entidades PROV



Fonte: Yolanda Gil e Miles (2013)

No PROV, *Entity* (Entidades) são tidas como qualquer coisa física, digital ou conceitual, a exemplo de uma página *Web*. “Os registros de proveniência podem descrever a proveniência das entidades, e a proveniência de uma entidade pode se referir a muitas outras entidades.” (GIL, Y; MILES, 2013, não paginado, tradução nossa).

Activity (Atividades) são definidas como ações e processos dinâmicos e são como entidades (*Entity*) que passam a existir quando seus atributos mudam para se tornarem novas entidades. Atividades podem gerar novas entidades, por exemplo, escrever um

documento traz o documento à existência, enquanto a revisão do documento traz uma nova versão à existência. (GIL, Y; MILES, 2013).

De acordo com Yolanda Gil e Miles (2013, não paginado, tradução nossa), “Um agente pode ser uma pessoa, um software, um objeto inanimado, uma organização ou outras entidades que podem ser responsabilizadas.” Um agente pode assumir uma responsabilidade em uma atividade que está ocorrendo, nesses casos considera-se que o agente foi associado com a atividade. Em alguns casos, um agente pode estar agindo em nome de agente, por exemplo, um funcionário pode agir em nome de uma organização, podendo ser expressa essa relação de responsabilidade na proveniência. (GIL, Y; MILES, 2013).

Outros conceitos fundamentais no PROV são: função, derivação, revisão, planos e tempo. A **função** específica o relacionamento entre uma entidade com uma atividade. “As funções também especificam como os agentes estão envolvidos em uma atividade, qualificando sua participação na atividade ou especificando para qual aspecto cada agente foi responsável.” (GIL, Y; MILES, 2013).

A **derivação** e a **revisão** estão relacionadas ao conteúdo, às características de uma entidade, então pode-se dizer que a primeira derivou da segunda. O PROV permite a descrição de alguns tipos de derivação, por exemplo, um documento pode passar por várias revisões ao longo do tempo. No PROV, o resultado de cada revisão é uma nova entidade e permite relacionar essas entidades fazendo uma descrição de que uma foi uma revisão de outra. (GIL, Y; MILES, 2013).

O PROV refere-se às **atividades** que seguem procedimentos predefinidos, como receitas, tutoriais, instruções ou fluxos de trabalho, como **planos**, e permite a descrição de que um plano foi seguido, por agentes, na execução de uma atividade. (GIL, Y; MILES, 2013).

O registro do **tempo** é fundamental para proveniência, nesse sentido, o PROV é capaz de registrar o cronograma de eventos da entidade e/ou da atividade. (GIL, Y; MILES, 2013).

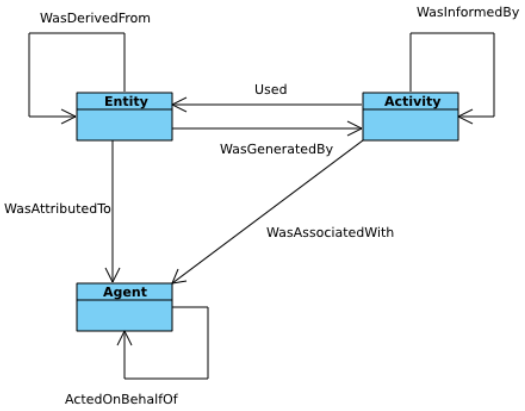
Em alguns casos, outras entidades alternativas e específicas podem ser requeridas, pois há diversas maneiras de descrever algo em um registro de proveniência, assim como, a necessidade de especificação de alguma entidade. (GIL, Yolanda; MILES, 2013).

3.4 The PROV Data Model (PROV-DM)

O PROV-DM é um modelo conceitual geral de dados, independente e extensível que possibilita a interoperabilidade da proveniência entre sistemas heterogêneos. Constitui a base para especificações da família PROV com estruturas centrais formando a essência de informações de procedência. O Modelo possui ainda estruturas estendidas que aprimoram e refinam as estruturas centrais para atender a usos mais avançados de proveniência. Está organizado em seis componentes: (1) entidades e atividades, e o momento em que eles foram criadas, usadas ou terminaram; (2) derivações de entidades; (3) agentes que carregam a responsabilidade das entidades que foram geradas e atividades que aconteceram; (4) um mecanismo de apoio proveniência para o empacotamento dos dados; (5) propriedades para vincular entidades que se referem à mesma coisa; (6) coleções formando uma estrutura lógica para os seus membros. (MOREAU; MISSIER, 2013a).

Na figura 6, é apresentada o primeiro componente da estrutura do PROV.

Figura 6 - Estruturas Essenciais do PROV



Fonte: Moreau e Missier, 2013a.

A estrutura base do PROV está modelada pelo diagrama *Unified Modeling Language* (UML)[3] e descreve o uso e a produção de entidades, atividades e agentes. De acordo com Moreau e Missier (2013a, não paginado, tradução nossa) “Uma entidade é um tipo físico, digital, conceitual ou outro tipo de coisa com alguns aspectos fixos; entidades podem ser reais ou imaginárias.” Exemplos de entidades podem ser consideradas, mas não se limitam a

3 UML é uma linguagem que auxilia na modelagem e na documentação de sistemas orientados a objetos.

um arquivo, uma ideia, um objeto. “Uma atividade é algo que ocorre durante um período de tempo e age sobre ou com entidades; pode incluir o consumo, processamento, transformação, modificação, realocação, uso ou geração de entidades.” (MOREAU; MISSIER, 2013a, não paginado, tradução nossa). Exemplos de atividades são: mover, copiar, duplicar entidades digitais, entre outros. Há uma forte ligação entre as entidades e as atividades, pois, as atividades utilizam entidades e atividades produzem entidades. Já o agente, refere-se a

> [...] algo que tem alguma forma de responsabilidade por uma atividade que ocorre, pela existência de uma entidade ou pela atividade de outro agente. Um agente pode ser um tipo particular de entidade ou atividade. Isso significa que o modelo pode ser usado para expressar a proveniência dos próprios agentes. (MOREAU; MISSIER, 2013a, não paginado, tradução nossa).

Confrme contsta na figura 6, as relações do modelos são: *Generation* (*WasGeneratedBy*); *Usage (Used); Communication (WasInformedBy*); *Derivation (WasDerivedFrom); Attribution (WasAttributedTo*); *Association* (*WasAssociatedWith*); e *Delegation (ActedOnBehalfOf)*.

O termo *Generation* (geração) refere-se à conclusão do ato de produzir, ou seja, a conclusão da produção de entidade. Já o termo *Usage* (uso) refere-se ao início do ato de utilizar as entidades, ou seja, é o começo de utilizar uma entidade por uma atividade. Se acordo com Moreau e Missier (2013a, não paginado, tradução nossa), há uma associação entre geração e uso, pois “A geração de uma entidade por uma atividade e seu uso subsequente por outra atividade é denominada comunicação.” O termo *Communication* (comunicação) corresponde a troca de alguma entidade por duas atividades, ou uma atividade usando alguma entidade gerada pela outra. Em alguns momentos, a utilização de uma entidade influencia a criação de outra, então é considerada como uma derivação (*derivation*), ou seja, a transformação de uma entidade em outra. (MOREAU; MISSIER, 2013a).

Outra relação é a atribuição (*Attribution*) de responsabilidade, ou seja, a atribuição de responsabilidade a um agente. Já a associação (*Association*), pode ser considerada como uma relação da atribuição de responsabilidade a um agente, indicando que o agente tinha uma função na atividade. Por fim, a Delegação (*Delegation*) pode ser definida como a atribuição de autoridade e responsabilidade a um agente para realizar uma atividade

específica. (MOREAU; MISSIER, 2013a).

Além da estrutura explicitada anteriormente, o PROV possui ainda, uma estrutura estendida que permite a inclusão de usos mais avançados da proveniência. Então são definidos mecanismos para as estruturas estendidas. De acordo com Moreau e Missier (2013a) as estruturas estendidas são definidas por uma variedade de mecanismos como subtipagem, relações expandidas, identificação opcional e novas relações. A proveniência da proveniência é retratada em um pacote *(bundle)* com um conjunto de dados que reflete a procedência e a credibilidade, permitindo assim, a garantia de acesso das informações de proveniência. Já a coleção é uma entidade que fornece uma estrutura para constituintes que devem ser entidades. (MOREAU; MISSIER, 2013a, não paginado, tradução nossa).

3.5 PROV-O

Editores responsáveis pela PROV-O: *The PROV Ontology* foram Timothy Lebo, da *Rensselaer Polytechnic Institute*, Satya Sahoo, da *Case Western Reserve University* e Deborah McGuinness, da *Rensselaer Polytechnic Institute*, dos Estados Unidos. "A Ontologia PROV (PROV-O) define a codificação da Linguagem de Ontologia da *Web* OWL2 do Modelo de Dados PROV (PROV-DM). (LEBO; SAHOO; MCGUINNESS, 2013). Dessa forma, O PROV-O estabelece "[...] um conjunto de classes, propriedades e restrições que podem ser usadas para representar e intercambiar informações de proveniência geradas em diferentes sistemas e sob diferentes contextos." (LEBO; SAHOO; MCGUINNESS, 2013).

No contexto do PROV-O, uma instância é um objeto individual em um domínio do discurso. Um conjunto de indivíduos que compartilha características comuns constitui uma classe. As propriedades são usadas para vincular indivíduos, classes ou criar uma hierarquia de propriedades.

> Os usuários do PROV-O podem precisar apenas usar partes de toda a ontologia, dependendo de suas necessidades e de quantos detalhes desejam incluir em suas informações de proveniência. Para isso, os termos PROV-O (classes e propriedades) são agrupados em três categorias para fornecer uma introdução incremental à ontologia: termos de ponto inicial, termos expandidos e termos para relacionamentos de qualificação. (LEBO; SAHOO; MCGUINNESS, 2013).

A Ontologia PROV, possui três classes principais, conforme originário do PROV-DM: entidade, atividade e agente. No quadro 4 são apresentadas as classes básicas da ontologia PROV.

Quadro 4 - Classes PROV-O

Classe	Definição
prov:Entity	Uma entidade é um tipo físico, digital, conceitual ou outro tipo de coisa com alguns aspectos fixos; entidades podem ser reais ou imaginárias.
prov:Activity	Uma atividade é algo que ocorre durante um período de tempo e age sobre, ou com entidades; pode incluir o consumo, processamento, transformação, modificação, realocação, uso ou geração de entidades.
prov:Agent	Um agente é algo que tem alguma forma de responsabilidade por uma atividade que ocorre, pela existência de uma entidade ou pela atividade de outro agente.

Fonte: Baseado em Lebo, Sahoo e Mcguinnness (2013).

O PROV-O apresenta nove propriedades: Geração; derivação; atribuição; início; uso; comunicação; fim; associação e delegação, conforme apresentado no quadro 5.

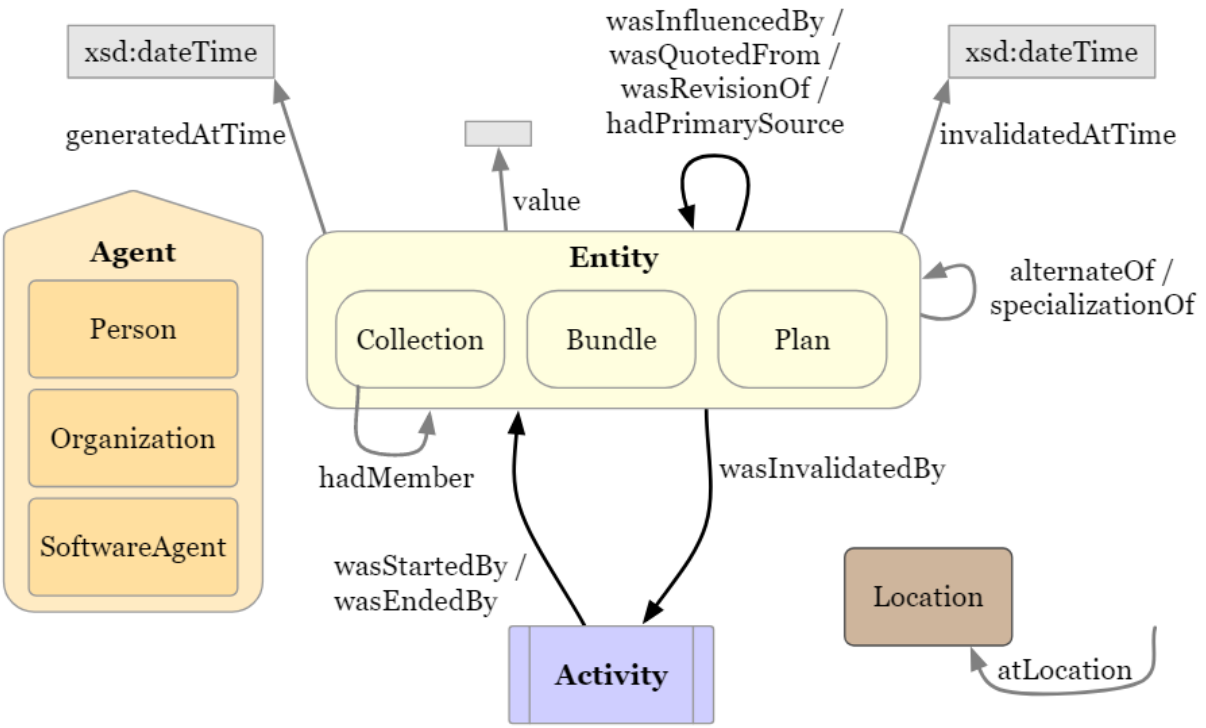
Quadro 5 - Propriedades do PROV-O

Propriedade	Definição
prov:wasGeneratedBy	Foi gerado pela conclusão da produção de uma nova entidade por uma atividade.
prov:wasDerivedFrom	Foi derivado de é uma transformação de uma entidade em outra, pode ser ainda uma atualização, ou a construção de uma nova entidade baseada em uma entidade pré-existente.
prov:wasAttributedTo	Foi atribuído para é a atribuição de uma entidade a um agente.
prov:startedAtTime	O início é quando uma atividade é iniciada por uma entidade, conhecida como acionador. Qualquer uso, geração ou invalidação envolvendo uma atividade segue o início da atividade.
prov:used	O uso é o começo de utilizar uma entidade por uma atividade.
prov:wasInformedBy	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
prov:endedAtTime	Fim é quando uma atividade é considerada encerrada por uma entidade. Qualquer uso, geração ou invalidação envolvendo uma atividade precede o final da atividade.
prov:wasAssociatedWith	Foi associado com é uma atribuição de responsabilidade a um agente de uma atividade.
prov:actedOnBehalfOf	Agiu em nome de ou Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que atua em nome de alguma responsabilidade pelo resultado do trabalho delegado.

Fonte: Baseado em Lebo, Sahoo e Mcguinnness (2013).

As classes e propriedades expandidas são complementares às classes principais e muitos termos são subclasses ou subpropriedades das classes principais, conforme apresentado na figura 7.

Figura 7 - Diagrama das classes principais



Fonte: Lebo, Sahoo e Mcguinnness (2013, não paginado).

As entidades estendidas auxiliam na explicitação das classes possibilitando diferenças entre elas. Nesse sentido são definidas sete subclasses: Coleção, coleção vazia, pacote, pessoas, agente de *software*, organização e local, conforme descrito no quadro 6.

Quadro 6 - Subclasses das propriedades estendidas

Classes	Definição
prov:Collection	Uma coleção é uma entidade que fornece uma estrutura para alguns constituintes, que são entidades deles. Estes constituintes são considerados membros das coleções.
prov:EmptyCollection	Uma coleção vazia é uma coleção sem membros.
prov:Bundle	Um pacote é um conjunto nomeado de descrições de proveniência, e é em si uma Entidade, permitindo assim que a proveniência da proveniência seja expressa.
prov:Person	Indivíduo ou identidade estabelecida por um indivíduo.
prov:SoftwareAgent	Um agente de software está executando o software.

prov:Organization	Uma organização é uma instituição social ou legal, como uma empresa, sociedade, etc.
prov:Location	Um local pode ser um local geográfico identificável (ISO 19112), mas também pode ser um local não geográfico, como um diretório, linha ou coluna.

Fonte: Baseado em Lebo, Sahoo e Mcguinnness (2013).

As propriedades expandidas possuem ainda 16 subpropriedades. Elas auxiliam na representação das subclasses no PROV-O. Elas são: *prov:alternateOf* (alternativa de), *prov:specializationOf* (especialização de), *prov:generatedAtTime* (geração no tempo), *prov:hadPrimarySource* (tem fonte primária), *prov:value* (valor), *prov:wasQuotedFrom* (foi citado de); *prov:wasRevisionOf* (foi revisão de); *prov:invalidatedAtTime* (invalidado na hora); *prov:wasInvalidatedBy* (foi invalidado por); *prov:hadMember* (tinha Membro); *prov:wasStartedBy* (foi qiniciado por); *prov:wasEndedBy* (foi finalizado); *prov:invalidated* (invalidado); *prov:influenced* (prov: influenciado); *prov:atLocation* (prov: at Location); *prov:generated* (prov: gerado), conforme apresentado no quadro 7.

Quadro 7 - Propriedades das classes expandidas

Propriedade	Descrição
prov: alternateOf	Duas entidades alternativas apresentam aspectos da mesma coisa. Esses aspectos podem ser iguais ou diferentes, e as entidades alternativas podem ou não se sobrepor no tempo.
prov: specializationOf	Uma entidade que é uma especialização de outra, compartilha todos os aspectos da última e, adicionalmente, apresenta aspectos mais específicos da mesma coisa que a segunda. [...] Exemplos de aspectos incluem um período de tempo, uma abstração e um contexto associado à entidade.
prov: generatedAtTime	Geração é a conclusão da produção de uma nova entidade por uma atividade.
prov: hadPrimarySource	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem o benefício da retrospectiva. [...] Como tal, é importante que fontes secundárias façam referência a essas fontes primárias das quais foram derivadas, para que sua confiabilidade possa ser investigada.
prov: value	Fornece um valor que é uma representação direta de uma entidade.
prov: wasQuotedFrom	Foi citado por é uma citação é a repetição de (algumas ou todas) de uma entidade, como texto ou imagem, por alguém que pode ou não ser seu autor original.
prov: wasRevisionOf	Foi revisado de é uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. Revisão é um caso particular de derivação.
prov: invalidatedAtTime	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade.
prov: wasInvalidatedBy	Foi invalidado por é Invalidação por meio da destruição, cessação ou expiração de

	uma entidade existente por uma atividade.
prov: hadMember	Uma coleção é uma entidade que fornece uma estrutura para alguns constituintes, que são entidades deles. Estes constituintes são considerados membros das coleções.
prov: wasStartedBy	Foi iniciado por é quando uma atividade é iniciada por uma entidade, conhecida como acionador.
prov: wasEndedBy	Foi encerrado por é quando uma atividade é encerrada por uma entidade.
prov:invalidated	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade.
prov:influenced	Influência é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
prov:atLocation	Um local pode ser um local geográfico identificável (ISO 19112), mas também pode ser um local não geográfico, como um diretório, linha ou coluna. Como tal, existem várias maneiras pelas quais a localização pode ser expressa, como por uma coordenada, endereço, marco e assim por diante.
prov:generated	Geração é a conclusão da produção de uma nova entidade por uma atividade.

Fonte: Baseado em Lebo, Sahoo e Mcguinnness (2013).

O PROV-O possui ainda, um conjunto de vinte (20) classes e propriedades que podem ser utilizadas para fornecer atributos adicionais às categorias das classes e propriedades básicas e estendidas. As classes definidas são: *prov:Influence* (influêncija); *prov:EntityInfluence* (Influência da Entidade); *prov:Usage* (Uso); *prov:Start* (Início); *prov:End* (Fim); *prov:Derivation* (Derivação); *prov:PrimarySource* (Fonte primária); *prov:Quotation* (Citação); *prov:Revision* (Revisão); *prov:ActivityInfluence* (Influência da atividade); *prov:Generation* (Geração); *prov:Communication* (Comunicação); *prov:Invalidation* (Invalidação); *prov:AgentInfluence* (Influência do agente); *prov:Attribution* (Atribuição); *prov:Association* (Associação); *prov:Plan* (Plano); *prov:Delegation* (Delegação); *prov:InstantaneousEvent* (Evento da instanciação); e *prov:Role* (Papel), conforme apresentado no quadro 8.

Quadro 8 - Classes PROV-O qualificado

Classe	Descrição
prov:Influence	Influência é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
prov:EntityInfluence	Entidade Influênciado é a capacidade de uma entidade ter um efeito sobre o caráter, desenvolvimento ou comportamento de outra por meio de uso, início, fim, derivação ou outro.

prov:Usage	O uso é o começo de utilizar uma entidade por uma atividade. Antes do uso, a atividade não havia começado a utilizar essa entidade e não poderia ter sido afetada pela entidade.
prov:Start	O início é quando uma atividade é considerada como iniciada por uma entidade, conhecida como acionador. [...] Qualquer uso, geração ou invalidação envolvendo uma atividade segue o início da atividade. Uma partida pode se referir a uma entidade desencadeadora que desencadeou a atividade ou a uma atividade, conhecida como acionador de partida, que gerou o acionador.
prov:End	Fim é quando uma atividade é considerada encerrada por uma entidade, conhecida como acionador. [...] Qualquer uso, geração ou invalidação envolvendo uma atividade precede o final da atividade. Um fim pode se referir a uma entidade desencadeadora que encerrou a atividade ou a uma atividade, conhecida como ender, que gerou o acionador.
prov:Derivation	Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova, ou a construção de uma nova entidade baseada em uma entidade pré-existente.
prov:PrimarySource	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem o benefício da retrospectiva. Por causa da objetividade das fontes primárias, elas "falam por si" de maneiras que não podem ser capturadas através do filtro de fontes secundárias. Como tal, é importante que fontes secundárias façam referência a essas fontes primárias das quais foram derivadas, para que sua confiabilidade possa ser investigada. Uma relação de fonte primária é um caso particular de derivação de materiais secundários de suas fontes primárias. Reconhece-se que a determinação de fontes primárias pode ser de interpretação e deve ser feita de acordo com as convenções aceitas no domínio do aplicativo.
prov:Quotation	Uma citação é a repetição de (algumas ou todas) de uma entidade, como texto ou imagem, por alguém que pode ou não ser seu autor original.
prov:Revision	Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. Revisão é um caso particular de derivação.
prov:ActivityInfluence	Influência da atividade é a capacidade de uma atividade de afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de geração, invalidação, comunicação ou outra.
prov:Generation	Geração é a conclusão da produção de uma nova entidade por uma atividade. Essa entidade não existia antes da geração e se torna disponível para uso após essa geração.
prov:Communication	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
prov:Invalidation	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade. A entidade não está mais disponível para uso (ou posterior invalidação) após a invalidação. Qualquer geração ou uso de uma entidade precede sua invalidação.
prov:AgentInfluence	AgentInfluence é a capacidade de um agente para influenciar o caráter, o desenvolvimento ou o comportamento de outro por meio de atribuição, associação, delegação ou outro.
prov:Attribution	Atribuição é a atribuição de uma entidade a um agente. Quando uma entidade e é atribuída ao agente ag, a entidade e foi gerada por alguma atividade não especificada que, por sua vez, foi associada ao agente ag. Assim, essa relação é útil

	quando a atividade não é conhecida ou irrelevante.
prov:Association	Uma associação de atividades é uma atribuição de responsabilidade a um agente de uma atividade, indicando que o agente tinha uma função na atividade. Além disso, permite que um plano seja especificado, que é o plano pretendido pelo agente para atingir algumas metas no contexto dessa atividade.
prov:Plan	Um plano é uma entidade que representa um conjunto de ações ou etapas pretendidas por um ou mais agentes para atingir alguns objetivos.
prov:Delegation	Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que atua em nome de retém alguma responsabilidade pelo resultado do trabalho delegado.
prov:InstantaneousEvent	O modelo de dados PROV é implicitamente baseado em uma noção de eventos instantâneos (ou apenas eventos), que marcam transições no mundo. Os eventos incluem geração, uso ou invalidação de entidades, bem como início ou término de atividades. Essa noção de evento não é de primeira classe no modelo de dados, mas é útil para explicar seus outros conceitos e sua semântica.
prov:Role	Uma função é a função de uma entidade ou agente em relação a uma atividade, no contexto de um uso, geração, invalidação, associação, início e fim.

Fonte: Baseado em Lebo, Sahoo e Mcguinnness (2013).

Foram definidas 25 propriedades no PROV-O. Elas são: *prov:wasInfluencedBy* (foi influenciado por); *prov:qualifiedInfluence* (Influência qualificada); *prov:qualifiedGeneration* (Geração qualificada); *prov:qualifiedDerivation* (Derivação qualificada); *prov:qualifiedPrimarySource* (fonte primária qualificada); *prov:qualifiedQuotation* (citação qualificada); *prov:qualifiedRevision* (revisão qualificada); *prov:qualifiedAttribution* (Atribuição qualificada); *prov:qualifiedInvalidation* (Invalidação qualificada); *prov:qualifiedStart* (Início qualificado); *prov:qualifiedUsage* (Uso qualificado); *prov:qualifiedCommunication* (Comunicação qualificada); *prov:qualifiedAssociation* (Associação qualificada); *prov:qualifiedEnd* (extremidade qualificada); *prov:qualifiedDelegation* (Delegação qualificada); *prov:influencer* (influenciador); *prov:entity* (entidade); *prov:hadUsage* (tinha uso); *prov:hadGeneration* (teve Geração); *prov:activity* (atividade); *prov:agent* (agente); *prov:hadPlan* (tinha plano); *prov:hadActivity* (Atividade); *prov:atTime* (no tempo); e *prov:hadRole* (tinha papel), conforme apresentado no quadro 9.

Quadro 9 - Propriedades do PROV-O qualificado

Propriedade	**Definição**
prov:wasInfluencedBy	Foi influenciado por é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de uso, início, fim, geração, invalidação, comunicação, derivação,

	atribuição, associação ou delegação.
prov:qualifiedInfluence	Influência é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
prov:qualifiedGeneration	Geração é a conclusão da produção de uma nova entidade por uma atividade. Essa entidade não existia antes da geração e se torna disponível para uso após essa geração.
prov:qualifiedDerivation	Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova, ou a construção de uma nova entidade baseada em uma entidade pré-existente.
prov:qualifiedPrimarySource	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem o benefício da retrospectiva. Por causa da objetividade das fontes primárias, elas "falam por si" de maneiras que não podem ser capturadas através do filtro de fontes secundárias. Como tal, é importante que fontes secundárias façam referência a essas fontes primárias das quais foram derivadas, para que sua confiabilidade possa ser investigada. Uma relação de fonte primária é um caso particular de derivação de materiais secundários de suas fontes primárias. Reconhece-se que a determinação de fontes primárias pode ser de interpretação e deve ser feita de acordo com as convenções aceitas no domínio do aplicativo.
prov:qualifiedQuotation	Uma citação é a repetição de (algumas ou todas) de uma entidade, como texto ou imagem, por alguém que pode ou não ser seu autor original. A citação é um caso particular de derivação.
prov:qualifiedRevision	Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. Revisão é um caso particular de derivação.
prov:qualifiedAttribution	Atribuição é a atribuição de uma entidade a um agente. Quando uma entidade e é atribuída ao agente ag, a entidade e foi gerada por alguma atividade não especificada que, por sua vez, foi associada ao agente ag. Assim, essa relação é útil quando a atividade não é conhecida ou irrelevante.
prov:qualifiedInvalidation	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade. A entidade não está mais disponível para uso (ou posterior invalidação) após a invalidação. Qualquer geração ou uso de uma entidade precede sua invalidação.
prov:qualifiedStart	O início é quando uma atividade é considerada como iniciada por uma entidade, conhecida como acionador. A atividade não existia antes do seu início. Qualquer uso, geração ou invalidação envolvendo uma atividade segue o início da atividade. Uma partida pode se referir a uma entidade desencadeadora que desencadeou a atividade ou a uma atividade, conhecida como acionador de partida, que gerou o acionador.
prov:qualifiedUsage	O uso é o começo de utilizar uma entidade por uma atividade. Antes do uso, a atividade não havia começado a utilizar essa entidade e não poderia ter sido afetada pela entidade.
prov:qualifiedCommunication	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
prov:qualifiedAssociation	Uma associação de atividades é uma atribuição de responsabilidade a um

	agente de uma atividade, indicando que o agente tinha uma função na atividade. Além disso, permite que um plano seja especificado, que é o plano pretendido pelo agente para atingir algumas metas no contexto dessa atividade.
prov:qualifiedEnd	Fim é quando uma atividade é considerada encerrada por uma entidade, conhecida como acionador. A atividade não existe mais após o seu término. Qualquer uso, geração ou invalidação envolvendo uma atividade precede o final da atividade. Um fim pode se referir a uma entidade desencadeadora que encerrou a atividade ou a uma atividade, conhecida como ender, que gerou o acionador.
prov:qualifiedDelegation	Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que atua em nome de retém alguma responsabilidade pelo resultado do trabalho delegado. Por exemplo, um estudante agia em nome de seu supervisor, que agia em nome do presidente do departamento, que agia em nome da universidade; todos esses agentes são responsáveis de alguma forma pela atividade que ocorreu, mas não dizemos explicitamente quem é o responsável e em que grau.
prov:influencer	Essa propriedade é usada como parte do padrão de influência qualificado. Subclasses de prov: Influenciar o uso dessas subpropriedades para referenciar o recurso (entidade, agente ou atividade) cuja influência está sendo qualificada.
prov:entity	A propriedade prov: entity faz referência a prov: entidade que influenciou um recurso. Esta propriedade aplica-se a uma prov: EntityInfluence, que é dada por uma subproperty de prov: qualifiedInfluence da prov influenciada: Entity, prov: Activity ou prov: Agent.
prov:hadUsage	O uso é o começo de utilizar uma entidade por uma atividade. Antes do uso, a atividade não havia começado a utilizar essa entidade e não poderia ter sido afetada pela entidade.
prov:hadGeneration	Geração é a conclusão da produção de uma nova entidade por uma atividade. Essa entidade não existia antes da geração e se torna disponível para uso após essa geração.
prov:activity	A propriedade prov: activity faz referência a prov: Atividade que influenciou um recurso. Esta propriedade aplica-se a uma prov: ActivityInfluence, que é dada por uma subproperty de prov: qualifiedInfluence da prov influenciada: Entity, prov: Activity ou prov: Agent.
prov:agent	A propriedade prov: agent faz referência a um prov: Agente que influenciou um recurso. Esta propriedade aplica-se a um prov: AgentInfluence, que é dada por uma subproperty de prov: qualifiedInfluence da prov influenciado: Entity, prov: Activity ou prov: Agent.
prov:hadPlan	Um plano é uma entidade que representa um conjunto de ações ou etapas pretendidas por um ou mais agentes para atingir alguns objetivos.
prov:hadActivity	Uma atividade é algo que ocorre durante um período de tempo e age sobre ou com entidades; pode incluir o consumo, processamento, transformação, modificação, realocação, uso ou geração de entidades.
prov:atTime	O modelo de dados PROV é implicitamente baseado em uma noção de eventos instantâneos (ou apenas eventos), que marcam transições no mundo. Os eventos incluem geração, uso ou invalidação de entidades, bem como início ou término de atividades. Essa noção de evento não é de primeira classe no

	modelo de dados, mas é útil para explicar seus outros conceitos e sua semântica.
prov:hadRole	Uma função é a função de uma entidade ou agente em relação a uma atividade, no contexto de um uso, geração, invalidação, associação, início e fim.

Fonte: Baseado em Lebo, Sahoo e Mcguinnness (2013).

Após o entendimento da ontologia PROV é possível fazer os relacionamentos e mapeamento dos elementos para os padrões utilizados no domínio bibliográfico. Entretanto, destaca-se que já nas definições das classes e das propriedade, há uma dificuldade de entendimento da semântica dos termos. Em muitos casos, as definições apresentadas são genéricas que se aplicam em algumas classes e propriedades gerais. Em outros casos, as definições se repetem causando confusão do uso dos termos. Apesar desses problemas, no próximo capítulo, buscou discutir e mapear os termos apresentados nas classes e propriedades do PROV para padrões no domínio bibliográfico.

4 METADADOS E SUAS TIPOLOGIAS

O processo de representação de recursos informacionais apresenta diversas características que corroboram com a identificação, a padronização, a qualidade e a integridade dos dados. Nesse contexto, os metadados têm papel fundamental para que o usuário possa identificar, localizar, recuperar e acessar um recurso. O termo metadados tem origem na Computação para descrever 'dados sobre dados' e segundo Vellucci (1998), o termo antecede a década de 60, mas começou a ser usado na década de 70, nos sistemas de bancos de dados. (HAYNES, 2018).

Ao longo dos anos, autores como Alves (2010); Alves e Santos (2013); Joudrey, Taylor e Wisser (2018); Méndez Rodríguez (2002); Pomerantz (2015); Zeng e Qin (2008, 2016), entre outros, têm discutido e estruturado definições mais contextualizadas do termo metadados na perspectiva da Ciência da Informação. De acordo com Joudrey, Taylor e Wisser (2018, p. 181, tradução nossa) "O que todos eles têm em comum é a noção de que os metadados são informações estruturadas que descrevem os atributos importantes dos recursos de informações para fins de identificação, descoberta, seleção, uso, acesso e gerenciamento." Nesse contexto, a tese de doutorado defendida por Alves (2010, p. 47-48) define os metadados 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.

Joudrey, Taylor e Wisser (2018, p. 181-182, tradução nossa) complementam que

> Os metadados podem incluir informações descritivas sobre o contexto, qualidade e condição, ou características dos dados. Esta definição implica que os metadados incluem não apenas informações descritivas, como aquelas encontradas em ferramentas de recuperação tradicionais para fins de descoberta de recursos, mas também informações necessárias para a gestão, uso e preservação do recurso de informação (por exemplo, dados sobre onde o recurso está localizado, como ele é exibido on-line, sua propriedade, sua condição).

Haynes (2018) apresenta diversos propósitos dos metadados como: identificar e descrever recursos; recuperar informação; gerenciar informações de recursos; gerenciar a

propriedade intelectual; dar suporte ao *e-commerce* e *e-government*; e ainda, a governança da informação.

O uso cada vez mais frequente dos metadados para as ações de representar, estruturar, gerenciar, preservar, usar e reusar informações, torna-se importante avaliar cada informação que será descrita, e ainda, pensar em outras informações de como o sistema irá funcionar, requisitos sobre *software* ou *hardware* em relação a preservação a longo prazo de recursos digitais, entre outros. Nesse contexto, destaca-se as categorias de cada tipo de metadados para seleção do padrão de metadados que deverá ser utilizado.

Ao categorizar os tipos de metadados, a literatura apresentou divergências principalmente no que diz respeito aos metadados administrativos. De acordo com Joudrey, Taylor e Wisser (2018), os metadados podem ser divididos em três grandes categorias, metadados administrativos, descritivos e estruturais. Entretanto não há um consenso e alguns autores (GILLILAND, 1999, 2008, 2016; MÉNDEZ RODRIGUEZ, 2002) identificaram cinco tipos de metadados, sendo eles administrativos, descritivos, de preservação, técnico e de uso.

A dificuldade em estabelecer as categorias dos metadados pode ser agravado, pois, alguns metadados com as mesmas características podem ser classificados em mais de uma categoria. A exemplo: data de criação pode estar na categoria descrição e administrativo, quando deseja fazer um levantamento da criação dos itens nos sistemas.

Joudrey, Taylor e Wisser (2018) apontam que já foram categorizados em onze (11) tipos de metadados, eles são: administrativo; comportamento; descritivo; avaliação de qualidade de imagem; meta-metadata; preservação; gerenciamento de registros ou manutenção de registros; direitos; estrutural; técnico; e rastreamento.

> Alguns autores ofereceram tipos adicionais de metadados a serem considerados, como metadados contextuais e metadados analíticos. Outros, no entanto, visualizam esses tipos adicionais como parte da categoria de metadados descritivos. Devemos lembrar que, como não há taxonomia formal de tipos de metadados com definições precisas, o que um autor pode se referir como metadados técnicos, outro pode chamar metadados estruturais ou metadados administrativos. As três categorias gerais empregadas neste texto refletem a categorização mais prevalente usada hoje. Alguns dos tipos mais frequentemente encontrados listados acima (técnicos, direitos, preservação e meta-metadata) são incluídos como subtipos de metadados administrativos. (JOUDREY; TAYLOR; WISSER, 2018, p. 184, tradução nossa)

Conforme apontado no quadro 10, de acordo com a literatura encontrada é possível observar as categorias definidas por cada autor. Os textos de Gilliland (1999, 2008, 2016) foram bases para diversos outros trabalhos, além de ser o primeiro texto localizado sobre as categorias dos metadados.

Quadro 10 - Quadro com as categorias de metadados

Autor	Baseado em	Categorias						
Gilliland (1999, 2008, 2016)	-	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Greenberg (2001)	-	Administrativo	Descoberta	-	-	Uso - Técnicos - Direitos	-	Autenticação
Méndez Rodríguez (2002)	-	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Senso e Rosa Piñero (2003)	Gilliland-Swetland (1999)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Riley (2004)	-	Administrativo - Tecnico - Direitos - Preservação	Descritivo	-	-	-	Estrutural	-
Haynes (2004)	Gilliland (1998)	Administrativo	Descritivo	Preservação	Técnicos	Uso		
Hillmann ; Marker; Brady (2008)	-	Administrativo	Descritivo	Preservação	-	Acesso/Uso	Estrutural	-
Alves (2010)	Gilliland-Swetland (1999), Senso e Rosa Piñero (2003) e Rosetto (2003)	Administrativo	Descritivo	Conservação	Técnicos	Uso	-	-
Miller (2011)	-	Administrativo - Técnico e preservação - Direitos - Uso	Descritivo	-	-	-	Estrutural	-
Alves e Santos (2013)	Gilliland (2008), Gilliland-Swetland (1999), Senso e Rosa Piñero (2003) e Rosetto (2003)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Pomerantz (2015)	-	Administrativo - Metadados técnicos - Estruturais - Proveniência - Preservação - Direitos - Meta-metadata	Descritivo	-	-	Uso - Data exhaust - Paradata	-	-
Zeng e Qin (2008; 2016)	Gilliland (2008)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Gartner (2016)	-	Administrativo - Tecnico - Direitos - Preservação	Descritivo	-	-	-	Estrutural	-

Riley (2017)	-	Administrativo - Tecnico - Direitos - Preservação	Descritivo	-	-	-	Estrutural	Linguagem de marcação
Haynes (2017)	Gilliland (2016)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Joudrey e Taylor (2018)	-	Administrativo - Tecnico, - Preservação - Direitos e acesso - Meta-metadata	Descritivo	-	-	-	Estrutural	-

Fonte: Elaborado pelo autor

A categorização dos metadados por administrativo, descritivo, preservação, técnicos e uso foi abordado por Gilliland (1999, 2008, 2016), Méndez Rodríguez (2002), Senso e Rosa Piñero (2003), Haynes (2004), Alves (2010), Alves e Santos (2013), Zeng e Qin (2008; 2016) e Haynes (2017), sendo que todos os textos têm como base os trabalhos de Gilliland (1999, 2008, 2016).

Autores como Riley (2004), Miller (2011), Gartner (2016) e Joudrey e Taylor (2018) definem três categorias, metadados administrativos, descritivos e estrutural. Após uma atualização, Riley (2017) coloca além dessas três categorias uma nova categoria: linguagem de marcação. Pomerantz (2015) coloca três categorias administrativo, descritivo e uso, sendo a categoria de metadados administrativos engloba os metadados técnicos, estruturais, proveniência, preservação, direitos, meta-metadata. Essa subcategorização de categorias nos metadados administrativos justifica-se segundo o autor, pois, muitos metadados possuem sobreposição de função, tendo como o objetivo principal de auxiliar no gerenciamento do recurso informacional.

Com intuito de esclarecer um direcionamento em relação aos metadados administrativos, objeto de pesquisa nesta tese, nas próximas seções serão discutidas as categorias e estabelecer as características dos metadados administrativos, descritivos, técnicos, uso, estruturais, proveniência, preservação, direitos, meta-metadata, estruturais, linguagem de marcação.

4.1 Metadados administrativos

A definição dos metadados administrativos, são similares em todos os autores, o que diferencia são as subcategorias que são usadas para dividir os metadados administrativos. A

definição de metadados administrativos, tem sido abordado como

> [...] criados para fins de gerenciamento, tomada de decisão e manutenção de registros. Eles fornecem informações sobre os requisitos técnicos, de preservação e armazenamento de recursos de informações, especialmente objetos digitais. Metadados administrativos auxiliam no monitoramento, acesso, reprodução, digitalização e backup de recursos digitais. (JOUDREY; TAYLOR; WISSER, 2018, p. 191-192).

Para os autores Joudrey, Taylor e Wisser (2018) os metadados administrativos têm como subcategorias os metadados Técnicos, Preservação, Direitos e acesso e meta-metadata. Alves e Santos (2013), apresenta uma definição próxima de Joudrey, Taylor e Wisser (2018), entretanto não subdivide os metadados administrativos. Para Alves e Santos (2013) os metadados administrativos

> [...] são metadados usados no gerenciamento e administração dos recursos de informação. Esse tipo de metadado fornece informações como: data de criação dos recursos, tipos de arquivos, formas de acesso, controle de diretos e reproduções, informação sobre registros legais, informação sobre localização etc.

Na mesma linha de Joudrey, Taylor e Wisser (2018) e Alves e Santos (2013), Miller (2011), discute que os metadados administrativos são usados para administrar e gerenciar coleções e objetos digitais. Para Miller (2011) são subcategorias dos metadados administrativos, os metadados ténicos, preservação, direitos e uso, conforme relatado na exemplificação a seguir:

> Exemplos de metadados administrativos são: o nome da instituição que cria os objetos digitais, a data da digitalização, o equipamento de digitalização usado, o nome do arquivo digital mestre, o arquivo de exibição e o arquivo de miniatura, dados do ciclo de vida da informação, como datas de criação de arquivos digitais, revisões subsequentes, tempo de revisão para retenção, arquivamento ou descarte, nomes criadores e revisores do recurso e níveis de autorização para uma determinada função. (MILLER, 2011, p. 12, tradução nossa).

Pomerantz (2015, p. 17-18, tradução nossa) defende que os “Metadados administrativos fornecem informações sobre a origem e a manutenção de um objeto: por exemplo, uma fotografia pode ter sido digitalizada usando um tipo específico de *scanner* em uma resolução específica e pode ter algumas restrições de direitos autorais associadas a ela.” Observa-se nesta definição do autor que da proveniência, metadados técnicos, estruturais, direitos, meta-metadata e preservação como subcategorias.

Gilliland (2008; 2016, não paginado, tradução nossa) define os metadados administrativos, como aqueles que são usados para

> [...] gerenciar e administrar coleções e recursos de informação. Informação de aquisição e avaliação; Direitos e acompanhamento de reprodução; Documentação de requisitos e protocolos legais, culturais e de acesso à comunidade; Informação de localização; Critérios de seleção para digitalização; Documentação de repatriamento digital, são exemplos de metadados administrativos.

Apesar da autora não subdividir os metadados administrativos pela definição apresentada, observa-se informações sobre direitos e acesso, além da questão do gerenciamento e administração de coleções.

Baseados em Gilliland (2008) e Riley (2004), Zeng e Qin (2016) definem que os metadados administrativos são "[...] usados no gerenciamento e administração de coleções e recursos de informações (exemplos incluem aquisição de informações, direitos e rastreamento de reprodução, requisitos legais de acesso e informações de localização)." O mesmo ocorre com a definição de Zeng e Qin (2008; 2016), apesar de não subdividir os metadados administrativos em subcategorias, sua definição apresenta a questão do gerenciamento da coleção, além de informações sobre direitos e acesso.

Segundo Riley (2017, p. 10, tradução nossa) "Metadados administrativos é um termo genérico referindo-se às informações necessárias para gerenciar um recurso ou relacionadas à sua criação." Greenberg (2001) faz diversos apontamentos sobre os metadados administrativos e inclui a proveniência, propriedade, diretos, preservação e meta-metadata como informações pertencentes à categoria, entretanto, não faz uma subdivisão.

> Os metadados de administração (ou administrativos) auxiliam na administração e no cuidado custodial de um objeto. Proveniência, data de aquisição, método de aquisição (por exemplo, compra / presente), restrições, propriedade e metadados da ação de preservação suportam atividades administrativas. Os metadados administrativos também podem incluir metametadata, que são metadados sobre os metadados (por exemplo, informações sobre quem criou os metadados). [...] Metadados administrativos são frequentemente, embora nem sempre, limitados da exibição pública. (GREENBERG, 2001, p. 919, tradução nossa).

A partir das definições propostas por Gilliland (2008, 2016) e Riley (2004; 2017), Zeng e Qin (2008, 2016), Pomerantz (2015), Alves e Santos (2013), Miller (2012), Joudrey, Taylor e Wisser (2018) o conceito do termo "metadados administrativos" são informações necessárias para administrar e gerenciar coleções, auxiliar na tomada de decisão e

manutenção de registros e dos recursos informacionais ou relacionadas à sua criação. Fornece informações sobre os requisitos técnicos, de preservação e de armazenamento de recursos de informações, a origem e a manutenção de um objeto e auxiliam no monitoramento, acesso, reprodução, digitalização e *backup* de recursos.

Os exemplos de metadados administrativos correspondem à: data de criação dos recursos, tipos de arquivos, formas de acesso, controle de diretos e reproduções, informação sobre registros legais, informação sobre localização, o nome da instituição que cria os objetos digitais, a data da digitalização, o equipamento de digitalização usado, o nome do arquivo digital mestre, o arquivo de exibição e o arquivo de miniatura, dados do ciclo de vida da informação, como datas de criação de arquivos digitais, revisões subsequentes, tempo de revisão para retenção, arquivamento ou descarte, nomes criadores e revisores do recurso e níveis de autorização para uma determinada função, informação de aquisição e avaliação; direitos e acompanhamento de reprodução; documentação de requisitos e protocolos legais, culturais e de acesso à comunidade; informação de localização; critérios de seleção para digitalização; documentação de repatriamento digital, são exemplos de metadados administrativos, aquisição de informações, direitos e rastreamento de reprodução, requisitos legais de acesso e informações de localização.

Para identificar os metadados que auxiliam na gestão de um sistema, é necessário entender as outras categorias de metadados. Segundo Joudrey, Taylor e Wisser (2018, p. 191-192), os metadados administrativos incluem informações como: requisitos de *software* e *hardware*; propriedade, direitos, permissões, acesso legal e informações sobre reprodução; uso da informação; característica de arquivos; controle de versão; informação de digitalização; dados de autenticação e segurança; informações de preservação.

De acordo com Joudrey; Taylor e Wisser (2018, p. 91), informações relativas aos requisitos de hardware e software correspondem a: *software* de criação, *hardware* de criação, informações de aquisição do sistema operacional: quando o recurso foi criado, modificado e/ou adquirido. Em relação à propriedade, direitos, permissões, acesso legal e informações sobre reprodução, estão relacionados aos materiais que são usados, quando, de que forma e por quem; usar estatísticas de rastreamento e circulação; rastreamento de usuários; reutilização de conteúdo; registros de exposições.

Dentre outras informações que podem ser consideradas nos metadados administrativos estão as informações de uso como: “[...] quais materiais são usados, quando,

de que forma e por quem; usar estatísticas de rastreamento e circulação; rastreamento de usuários; reutilização de conteúdo; registros de exposições [...]". (JOUDREY; TAYLOR; WISSER, 2018, p. 91, tradução nossa).

Informações características dos arquivos como tamanho, formato, regras de apresentação, informações de sequenciamento, tempo de execução, informações de compactação de arquivos e relativos ao controle de versão estão relacionadas à quais versões existem e qual status do recurso descrito; há formatos digitais alternativos, como *Hypertext Markup Language* (HTML) ou *Portable Document Format* (PDF) para texto, e *Graphics Interchange Format* (GIF) ou *Joint Photographic Experts Group* (JPG) para imagens. Informações de digitalização como taxas de compactação, proporções de escala, data de varredura, resolução, assim como, dados de autenticação e segurança como informações sobre inibidor, criptografia e senha, podem ser classificados como metadados administrativos. Informações de preservação como: informações de integridade, condição física, ações de preservação, atualização de dados, migração de dados, conservação ou reparo de artefatos físicos também fazem parte dos metadados administrativos conforme destacado por Joudrey, Taylor e Wisser (2018).

De acordo com Joudrey, Taylor e Wisser (2018, p. 91) os metadados administrativos suportam o gerenciamento, tomada de decisão, preservação, gerenciamento de cotas, suporte técnico, gerenciamento de metadados, rastreamento de uso, avaliação e manutenção de registros. São considerados subcategorias dos metadados administrativos: avaliação da qualidade de imagem; meta-metadata; preservação; manutenção de registros; direitos; técnico; rastreamento; uso. São elementos característicos, informações sobre: aquisição; condição; informação de digitalização; tamanho do arquivo; dados de criação de metadados; ações de reserva; responsabilidade; requisitos de *software*; requisitos de armazenamento; informações de uso; rastreamento do usuário.

> Ao contrário da maioria dos metadados descritivos, os metadados administrativos não são padronizados. A partir deste momento, não existe um único esquema administrativo de metadados que tenha sido adotado amplamente entre as instituições. Os metadados capturados para fins de gerenciamento, como quem toma decisões e suas informações de contato, tendem a ser específicos do repositório e são armazenados em uma variedade de formas, em vários lugares (por exemplo, metadados administrativos podem ou não ser incorporados a um registro compreendendo metadados descritivos primariamente). (JOUDREY; TAYLOR; WISSER, 2018, p. 91).

Conforme destacado por Joudrey, Taylor e Wisser (2018), a falta de definição da padronização dos tipos de metadados administrativos dificulta na utilização de um único padrão para os metadados administrativos, logo, alguns esquemas foram criados para abordar subtipos específicos de metadados como de preservação, técnicos, entre outros. Nesse sentido, serão apresentadas as demais tipologias localizadas sobre os metadados.

4.2 Metadados de autenticação

Alguns autores como Gilliland (2016) e Zeng e Qin (2016), abordam metadados de autenticação relacionados aos metadados administrativos, mas não colocam como uma subcategoria.

Já Greenberg (2001) destaca que os metadados de autenticação podem ocupar uma nova categoria. Segundo a autora, os metadados de autenticação "[...] suportam a avaliação da integridade de um objeto de informação, legitimidade e qualidade geral genuína." (GREENBERG, 2001, p. 919, tradução nossa). Exemplos de metadados de autenticação são: fonte, relação, versão / edição e assinatura digital são exemplos de metadados que ajudam a determinar a autenticidade de um objeto de informação.

4.3 Metadados de preservação

Autores como Riley (2004), Miller (2011), Pomerantz (2015), Gartner (2016), Riley (2017), Joudrey; Taylor e Wisser (2018) colocam os metadados de preservação como subcategoria dos metadados administrativos, outros autores como Gilliland (1999, 2008, 2016), Méndez Rodríguez (2002), Senso e Rosa Piñero (2003), Haynes (2004; 2017), Alves (2010), Alves e Santos (2013), Zeng e Qin (2008; 2016) colocam os metadados de preservação como categoria independente.

Segundo Alves e Santos (2013), os metadados de preservação estão

> [...] relacionados com a conservação e a preservação dos recursos de informação. Esse tipo fornece informações sobre as condições físicas de um recurso, informações sobre as ações tomadas para conservar e preservar as versões físicas e digitais de um recurso etc.

Assim como Alves e Santos (2013), Zeng e Qin (2016,) apontam que os metadados de preservação estão atrelados à questão da gestão da preservação, das ações e das alterações

que ocorrem durante o ciclo de vida do recurso informacional. Exemplos de informações "[...] incluem documentação da condição física de recursos, ações tomadas para preservar versões físicas e digitais de recursos (por /exemplo, atualização de dados e migração) e alterações que ocorrem durante a digitalização ou preservação." (ZENG; QIN, 2016, p. 18, tradução nossa).

Para Gilliland (2016), os metadados de preservação estão relacionados à gestão de preservação de coleções e recursos de informação. A exemplo de informações sobre metadados de preservação, estão: "Documentação de condição física de recursos; Documentação de ações tomadas para preservar versões físicas e digitais de recursos (por exemplo, atualização de dados e migração); Documentação de quaisquer alterações ocorridas durante a digitalização ou preservação." (GILLILAND, 2016, não paginado, tradução nossa). Riley (2017) complementa que os metadados de preservação estão relacionados ao gerenciamento de longo prazo de arquivos.

Joudrey, Taylor e Wisser (2018, p. 193, tradução nossa) esclarecem que os

> Metadados de preservação são as informações necessárias para garantir o armazenamento a longo prazo e a usabilidade do conteúdo digital. Pode incluir informações sobre reformatação, migração, emulação, conservação, integridade de arquivos e proveniência. Elementos típicos de metadados de preservação podem incluir o seguinte: Identificadores; tipos estruturais; descrições de arquivo; Tamanhos; propriedades; ambientes de software e hardware; informação da fonte; história do objeto; história de transformação; assinaturas digitais de informação de contexto; verificação.

Diante do cenário da necessidade de preservação dos recursos informacionais, a *Library do Congress* dos Estados Unidos (LC), a *Online Computer Library Center* (OCLC), o *Research Libraries Group* (RLG) e a *National Library* da Austrália, entre outros, lançaram iniciativas para o estabelecimento de metadados de preservação, entretanto, não é o foco deste trabalho abordar essas iniciativas.

4.4 Metadados técnicos

Há uma grande dificuldade na literatura em separar as informações dos metadados técnicos das informações de preservação, uma vez que as duas estão diretamente relacionadas e quase indissociáveis. Joudrey, Taylor e Wisser (2018, p. 192, tradução nossa)

buscam contextualizar os metadados técnicos fazendo as seguintes questões: “Se uma instituição da informação recebe um recurso digital, eles saberiam o que é? Eles entenderiam o que isso pode fazer? Eles poderiam interagir com o ir ou fazer funcionar? Sem metadados técnicos, eles podem não conseguir usar o objeto.”

Autores como Gilliland (1999, 2008, 2016), Méndez Rodríguez (2002), Senso e Rosa Piñero (2003), Haynes (2004), Alves (2010), Alves e Santos (2013), Zeng e Qin (2008; 2016), Haynes (2017) categorizam os metadados técnicos como categoria independente. Para Riley (2004; 2017), Miller (2011), Pomerantz (2015), Gartner (2016), Joudrey; Taylor e Wisser (2018), os metadados técnicos pertencem à categoria administrativa.

De acordo com Gilliland (2016, não paginado), os metadados técnicos estão relacionados em como um sistema funciona ou como os metadados se comportam. Estão relacionados com os metadados técnicos informações sobre a documentação do *hardware* e *software*, informações sobre a digitalização de um recurso, tempo de resposta de um sistema e autenticação e segurança de informações de um sistema.

De acordo com Pomerantz (2015, p. 95, tradução nossa) os “Metadados técnicos fornecem informações sobre como um sistema funciona ou detalhes no nível do sistema sobre recursos.” Riley (2017) complementa que os metadados técnicos servem para decodificar e renderizar arquivos. Podem ser considerados exemplos de metadados técnicos, segundo Gilliland (2016), a Documentação de *hardware* e *software*; Informações processuais geradas pelo sistema (por exemplo, metadados de roteamento e eventos); Informação técnica de digitalização (por exemplo, formatos, taxas de compressão, rotinas de escalonamento); Acompanhamento dos tempos de resposta do sistema; Dados de autenticação e segurança (por exemplo, chaves de criptografia, senhas).

Na mesma perspectiva, Zeng e Qin (2016) definem os metadados técnicos baseados em Gilliland (2008) e Riley (2004) que precisa registrar

> [...] como um sistema funciona ou metadados se comportam (exemplos incluem informações sobre requisitos de *hardware* e *software*), digitalização técnica (como formato, taxas de compactação e rotinas de dimensionamento) e dados de autenticação e segurança (por exemplo, chaves de criptografia e senhas). (ZENG; QIN, 2016, p. 19, tradução nossa).

Paralela à definição de Zeng e Qin (2008; 2016), Alves e Santos (2013) definem os metadados técnicos como que estão “[...] relacionados com o funcionamento dos sistemas e o comportamento dos metadados. Esse tipo fornece informações sobre *hardware* e

software, digitalização, controle do tempo de resposta dos sistemas, autenticidade e segurança dos dados (criptografia e senhas) etc.”

De acordo com o dicionário do PREMIS, apresenta uma intersecção entre os metadados técnicos e a preservação.

> Metadados técnicos descrevem as características físicas e não intelectuais dos objetos digitais. Metadados técnicos detalhados e específicos de formato são claramente necessários para implementar a maioria das estratégias de preservação, mas o grupo não tinha tempo nem conhecimento para lidar com metadados técnicos específicos de formato para vários tipos de arquivos digitais. Portanto, restringiu os metadados técnicos incluídos no Dicionário de Dados às unidades semânticas que eles acreditavam aplicar a objetos em todos os formatos. O desenvolvimento adicional de metadados técnicos é deixado para os especialistas formatarem.” (PREMIS EDITORIAL COMMITTEE, 2015, p. 32, tradução nossa)

Nesse contexto, Joudrey, Taylor e Wisser (2018) ressaltam a importância dos metadados técnicos para entender a natureza do recurso, os ambientes de *software* e *hardware* que o recurso foi criado e o que é necessário para tornar o recurso acessível aos usuários.

> Os metadados técnicos descrevem as características, origens e ciclos de vida de documentos digitais e são essenciais para a preservação do recurso para uso futuro. Como os metadados técnicos são específicos do formato, diferentes esquemas são usados para diferentes tipos de recursos (por exemplo, um fluxo de vídeo requer metadados técnicos diferentes de uma imagem digital porque eles funcionam de maneiras diferentes e têm características diferentes). (JOUDREY; TAYLOR; WISSER, 2018, p. 191-192, tradução nossa).

Miller (2011) busca estabelecer uma relação entre metadados técnicos e de preservação. Apesar de deixar claro que possuem características distintas, muitas informações são usadas para ambos os tipos. A exemplo disso, são destacadas

> Informações necessárias para a preservação a longo prazo do objeto digital, migração para outros formatos digitais como software e hardware mudam ao longo do tempo. Por exemplo: tipo de scanner usado, resolução de digitalização original, especificações de edição de imagem. (MILLER, 2012, p. 12, tradução nossa).

Por outro lado, há grande relacionamento entre os metadados técnicos e os metadados estruturais, mas vale destacar que não são a mesma coisa. “A maior diferença entre os pacotes é que os metadados estruturais são primariamente legíveis por máquina e

processáveis por máquina, enquanto as informações técnicas são para os humanos lerem e entenderem.” (JOUDREY; TAYLOR; WISSER, 2018, p. 191-192, tradução nossa).

4.5 Meta-metadata

A categoria meta-metadata é incorporada aos metadados administrativos por Pomerantz (2015) e Joudrey, Taylor e Wisser (2018). De acordo com os autores, os meta-metadata são informações sobre os metadados do recurso descrito.

> [...] meta-metadata preenche a função administrativa das próprias descrições de metadados: uma descrição de meta-metadata fornece informações sobre quando e como uma descrição de metadados foi criada (por quem, de onde e com quais padrões), que detalhes técnicos contém e quem tem privilégios de acesso aos metadados armazenados. (ZENG; QIN, 2016, p. 20-21, tradução nossa)

Paralelo à definição de Zeng e Qin (2016), Joudrey, Taylor e Wisser (2018, p. 196, tradução nossa), explicam que os meta-metadata podem rastrear dados administrativos de um recurso e seus metadados.

> Os meta-metadata são importantes para garantir a autenticidade dos metadados dos metadados e acompanhar o processo interno. Embora alguns meta-metadata residam em alguns tipos de registros descritivos (e.x., as informações de criação de registros e as datas de modificação em um registro MARC), outros meta-metadata devem ser rastreados de outras maneiras.

No ano de 2003, houve um grupo de trabalho da *Dublin Core Metadata Initiative* (DCMI) com o intuito de criar um padrão de meta-metadata. Nesse contexto, Hytte Hansen e Leif Andresen criaram o “AC- *Administrative Components*”, um conjunto de elementos para descrever e gerenciar metadados. Seus elementos incluem informações sobre: Manipulação; Ação; Base de dados; Contato; Afiliação; Transmissor. Entretanto, aparentemente não houve continuidade da proposta. (JOUDREY; TAYLOR; WISSER, 2018).

4.6 Metadados de Proveniência

Com a facilidade de copiar informações digitais, os metadados de proveniência tornam-se importantes na medida em que asseguram a fonte original que a informação está sendo extraída. De acordo com Yolanda Gil et al. (2010, não paginado, tradução nossa), a proveniência é "[...] um registro que descreve entidades e processos envolvidos na produção e entrega ou de outra forma influenciando esse recurso." De acordo com Pomerantz (2015, p. 101, tradução nossa) a "[...] proveniência significa não apenas a história de um recurso, mas as relações entre esse recurso e outras entidades que influenciaram sua história."

> A proveniência de metadados descreve e acompanha os agentes responsáveis, influenciando ações, eventos associados que causaram alterações nos metadados. Histórico de alterações de um esquema de metadados usado em um serviço é crucial para acompanhar as alterações nas instâncias de metadados criadas com base esquema. Portanto, a proveniência de um esquema de metadados é crucial para manter os metadados corretamente e consistentemente interpretável e pode incluir histórico de alterações do esquema, bem como relacionamentos para outras entidades, como padrões básicos e requisitos do sistema. (LI; SUGIMOTO, 2014, p. 149, tradução nossa).

Nesse contexto, segundo Pomerantz (2015, p. 102-103, tradução nossa), "Metadados de proveniência é um mecanismo para fornecer dados sobre entidades e seus relacionamentos com o recurso e com outras entidades." Segundo o autor, diversos padrões podem ser utilizados para representar a proveniência.

> Diversos esquemas de metadados de procedência existem atualmente; a padronização que ocorreu em outros domínios e para outros casos de uso (Dublin Core para uso geral, Getty thesauri para objetos de arte, Exif para imagens digitais, etc.) ainda tem que emergir para proveniência. Esses esquemas de proveniência compartilham muitas características: todos eles são compostos de conjuntos de elementos que identificam características do recurso ou de entidades que o influenciaram, e todos categorizam relacionamentos entre recursos e entidades. (POMERANTZ, 2015, p. 103, tradução nossa).

Entretanto, o entendimento dos metadados de proveniência ainda possuem lacunas e necessitam de estudos aprofundados.

4.7 Metadados descritivos

Entre as categorias de metadados, os metadados descritivos são os mais discutidos e melhor estabelecidos na literatura. De acordo com Alves e Santos (2013) os metadados

descritivos são usados para "[...] descrever, identificar e representar recursos de informações. Esse tipo fornece informações relacionadas com a catalogação, como título, autor, imprenta, data, resumo, palavras-chave, e ainda a relação dos *hiperlinks* entre os recursos, anotações de usuários etc." Gilliland (2016, não paginado, tradução nossa) esclarece que os metadados descritivos são usados para

> [...] identificar, autenticar e descrever coleções e recursos de informações confiáveis relacionados. Metadados gerados pelo criador e sistema original; Pacote de informação de submissão; Registros de catalogação; Encontrar ajudas; Controle de versão; Índices especializados; Informação curatorial Relações vinculadas entre recursos; Descrições, anotações e emendas de criadores e outros usuários.

No ponto de vista de Joudrey, Taylor e Wisser (2018, p. 184, tradução nossa), os metadados descritivos, "Descrevem as características identificadoras e os contextos intelectuais dos recursos informacionais para fins de descoberta, identificação, seleção, aquisição, contexto e compreensão."

Para Zeng e Qin (2016), baseado em Gilliland (2008) e Riley (2004), os metadados descritivos são usados para: "[...] identificar e descrever coleções e recursos de informações relacionadas (exemplos incluem registros de catalogação, ferramentas de localização, índices especializados e informações curatoriais)."

4.8 Metadados de direitos, acesso e uso

A literatura apresenta divergência na denominação da categoria direitos e acesso com metadados de uso. Autores como Riley (2004; 2017), Miller (2011), Pomerantz (2015), Gartner (2016), Joudrey, Taylor e Wisser (2018) utilizam o termo direitos como subcategoria dos metadados administrativos, apenas Joudrey, Taylor e Wisser (2018) que definem a categoria como direitos e acesso. Outros autores como Gilliland (1999, 2008, 2016), Méndez Rodríguez (2002), Senso e Rosa Piñero (2003), Haynes (2004, 2017), Alves (2010), Alves e Santos (2013), Pomerantz (2015), Zeng e Qin (2008; 2016) utilizam a denominação uso e categorizam como uma categoria independente. Miller (2011) categoriza os metadados de uso e direito como subcategorias dos metadados administrativos e Pomerantz (2015) categoriza os metadados de direitos como subcategoria de metadados administrativos e a categoria metadados de uso como independente.

Miller (2011, p. 12, tradução nossa) expõe que os metadados de direitos são

> Informações sobre propriedade, direitos autorais, restrições de uso e reprodução. Por exemplo, uma declaração de direitos autorais, informações sobre restrições de uso e reprodução de uma imagem digital, restrições de acesso se limitadas a apenas determinados usuários, método de pagamento para comprar ou baixar uma imagem de resolução total.

Complementar à definição de Miller (2011), Riley (2017, não paginado, tradução nossa) esclarece que os metadados de direitos estão relacionados aos direitos de propriedade intelectual associados ao conteúdo.

> Direitos e metadados de acesso são informações sobre quem tem acesso aos recursos de informações, quem pode usá-los e com quais objetivos. Ele lida com questões de direitos de propriedade intelectual dos criadores e os acordos legais que permitem aos usuários acessar essas informações. Aborda questões como: Quem pode acessar um recurso de informações e com quais objetivos? Quem pode fazer cópias? Quem possui o material? Existem diferentes categorias de objetos de informação na coleção? Existem diferentes categorias de usuários que podem acessar diferentes combinações desses objetos? Em direitos e metadados de acesso, informações sobre partes, conteúdos e transações podem ser encontradas. Alguns elementos de metadados de direitos típicos incluem o seguinte: Categorias de acesso; Identificadores; Nomes de criadores; Nomes de titulares de meias; Datas da criação; Status de direitos autorais; Termos e Condições; Restrições de acesso; Períodos de disponibilidade; Informações de uso; Opções de pagamento. (JOUDREY; TAYLOR; WISSER, 2018, p. 193, tradução nossa)

Muitas vezes metadados de direitos são colocados como similares aos metadados de uso, pois algumas informações possuem dupla função conforme destacado por Miller (2012, p. 12, tradução nossa), os metadados de uso "[...] podem ser separados dos metadados de direitos, mas também podem se sobrepor a eles. Por exemplo: dados sobre o número de vezes que uma imagem foi visualizada."

De acordo com Alves e Santos (2013), os metadados de uso são "[...] relacionados com o nível e tipo de uso dos recursos de informação. Esse tipo fornece informações sobre os registros de exibição, controle de uso e usuários, controles de acesso, informação sobre versões múltiplas, logs de acesso etc." Riley (2004), Gilliland (2008) e Zeng e Qin (2016, p. 19, tradução nossa) estabelecem os metadados de uso "[...] relacionados ao nível e tipos de uso de coleções e recursos de informação (exemplos incluem registros de circulação, registros de exibições físicas e digitais, uso, reutilização, pesquisa e rastreamento de usuários)."

Greenberg (2001, p. 919, tradução nossa) apresenta uma forma diferente de

entender os metadados de uso. Em sua concepção, os metadados de uso incluem informações técnicas e intelectuais para que o usuário possa utilizar os recursos informacionais procurados.

> Metadados de uso permitem a exploração técnica e intelectual de um objeto de informação. A exploração técnica inclui requisitos do sistema, formato, localização (por exemplo, endereço físico ou virtual) e outros metadados que afetam o acesso a objetos para o computador ou para um indivíduo. A exploração intelectual inclui direitos de propriedade, restrições de políticas e outros metadados de termos e condições para replicação de conteúdo e citações de publicação. Juntos, os metadados técnicos e intelectuais determinam quem pode, o que (por exemplo, uma parte de um objeto), onde, quando e como usar objetos. A classe se sobrepõe ao que foi identificado como metadados estruturais.

Segundo Pomerantz (2015, p. 18, tradução nossa), "[...] metadados de uso fornecem informações sobre como um objeto foi usado: por exemplo, o editor de um livro eletrônico pode rastrear quantos downloads o livro recebeu, em que datas e dados de perfil sobre os usuários que fizeram o download." Segundo o autor, os metadados de uso podem gerar diversas novas informações como "*Data Exhaust*". *Data Exhaust* corresponde às informações de navegação do usuário, ou seja, as trilhas dos dados do usuário ao usar a internet. Outro termo utilizado por Pomerantz (2015) é "*Paradata*" que refere aos dados sobre o uso dos objetos de aprendizagem digital pelos usuários adotado pela *National Science Digital Library* (NSDL). Entretanto, vale destacar que esses termos "*Data Exhaust*" e "*Paradata*" ainda não são consenso na literatura, sendo "*Paradata*" utilizado ainda como sinônimo para "*metametadata*".

Para Gilliland (2016, não paginado, tradução nossa) os metadados de uso estão relacionados "[...] ao nível e tipo de uso de coleções e recursos de informação. Registros de circulação; Registros de exposições físicas e digitais; uso e rastreamento de usuários; Reutilização de conteúdo e informações sobre multiversão; Registros de pesquisa; Metadados de direitos."

Apesar de alguns autores incluir informações de uso e direitos na mesma categoria, essas informações poderiam ser separadas pois possuem finalidades diferentes. Informações de direitos irão auxiliar no controle de informações como, quem poderá ter acesso e como as informações poderão ser utilizadas. Já as informações de uso, irão auxiliar em informações de como o objeto foi usado, proporcionando entre outras informações, estatísticas de acesso e utilização.

4.9 Metadados estruturais

De acordo com Joudrey, Taylor e Wisser (2018) o uso de metadados estruturais não é novo, mas a terminologia usada para descrevê-lo é mais recente. Para tanto, apenas Riley (2004, 2017), Miller (2011), Gartner (2016), Joudrey, Taylor e Wisser (2018) definem os metadados estruturais como uma categoria de metadados, já Pomerantz (2015) inclui como subcategoria dos metadados administrativos. Nesse contexto, Joudrey, Taylor e Wisser (2018, p. 193, tradução nossa) definem que

> Metadados estruturais referem-se à composição ou organização interna do objeto digital, conjunto de dados ou outro recurso de informações que está sendo descrito. São os dados necessários para garantir que um recurso digital funcione corretamente e possa ser usado e navegado pelo usuário.

Pomerantz (2015, p. 17-18, tradução nossa) complementa que "Metadados estruturais fornecem informações sobre como um objeto é organizado: por exemplo, um livro é composto de capítulos, um capítulo é composto de páginas e esses capítulos e páginas devem ser reunidos em uma ordem específica."

De acordo com Miller (2011, p. 12, tradução nossa), os metadados estruturais são "[...] usados para estruturar internamente um objeto digital complexo de fornecer estrutura para relações entre um conjunto de objetos digitais intimamente relacionados. Por exemplo: um único livro digitalizado como vários arquivos de imagem (objeto digital complexo)."

Joudrey, Taylor e Wisser (2018, p. 193, tradução nossa) explicam que os metadados estruturais "Referem-se a como os arquivos relacionados individuais são unidos para criar um objeto digital funcional, como o objeto pode ser exibido em diversos sistemas e como ele pode ser armazenado e divulgado. Ele lida com o que é o recurso, o que ele faz e como funciona." Segundo Joudrey, Taylor e Wisser (2018), os metadados estruturais abrangem os seguintes tipos de informações: tipos de documentos e sua estrutura; tipos de arquivo; comportamentos ou funcionalidade de objetos; protocolos de pesquisa associados; relações hierárquicas; sequenciamento e agrupamento de arquivos; objetos; informação de paginação; arquivos associados.

Apesar de importantes para estruturação de um recurso fragmentado, ou mesmo,

auxiliando os metadados de preservação para recomposição do recurso informacional, os metadados estruturais nem sempre são representados e usados adequadamente.

> Metadados estruturais podem ser incluídos nos cabeçalhos ou corpos de alguns tipos de documentos eletrônicos, mas na maioria dos esquemas de metadados, os elementos estruturais não são bem representados. [...] Às vezes, os metadados estruturais são chamados de metadados de exibição e, às vezes, são confundidos com metadados técnicos [...](JOUDREY; TAYLOR; WISSER, 2018, p. 193, tradução nossa).

De acordo com Riley (2017, p. 10, tradução nossa), os metadados estruturais fazem os "Relacionamentos de partes de recursos entre si. Linguagens de marcação; integra metadados e sinalizadores para outros recursos estruturais ou semânticos no conteúdo." Segundo Joudrey, Taylor e Wisser (2018, p. 193, tradução nossa), os metadados estruturais apresentam "[...] informações técnicas necessárias para garantir que um recurso digital funcione corretamente, sejam exibidas na tela e possam ser navegadas pelos usuários."

Pode-se considerar como metadados estruturais, informações sobre: Comportamento; Sequenciamento de arquivos; próxima página; Página anterior; Mapa de recursos. (JOUDREY; TAYLOR; WISSER, 2018). Alguns exemplos de padrões de metadados estruturais são: *Material eXchange Format* (MXF); *Metadata Encoding & Transmission Standard* (METS); *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE); *Page-turner models*.

Apesar de muitos autores colocarem os metadados estruturais como uma categoria independente, Pomerantz (2015) coloca que os metadados estruturais também são subcategorias dos metadados administrativos. "Os metadados estruturais e de preservação são às vezes considerados subcategorias de metadados administrativos, pois os dados sobre a estrutura de um objeto e como preservá-lo são necessários para administrar o objeto." (POMERANTZ, 2015, p. 17-18, tradução nossa).

4.10 *Markup languages*

Apenas Riley (2017) coloca as linguagens de marcação como categoria de metadados e há poucas informações sobre sua definição e uso. As linguagens de marcação "Integram metadados e sinalizadores para outros recursos estruturais ou semânticos no conteúdo."

(RILEY, 2017, não paginado, tradução nossa). De acordo com a autora, as linguagens de marcação misturam metadados e conteúdo e são importantes para destacar determinados elementos de um recurso.

A categorização das linguagens de marcação como tipologia de metadados, está relacionada que as linguens de marcação são metadados simples.

Dessa forma, para melhor elucidar e contextualizar os metadados de proveniência, entre as tipologias de metadados, foi necessário estabelecer uma organização das tipologias dos metadados, pois além dos metadados de proveniência, há diversas outras tipologias que surgiram que ainda pouco foram estudadas, como meta-metadata, metadados de autenticação, linguagens de marcação entre outras.

Assim, após a revisão de literatura sobre as tipologias de metadados, foi possível propor um novo conjunto de definições e hierarquia baseadas nos autores descritos neste capítulo, conforme apresentado no quadro 11.

Quadro 11 - Definição dos tipos de metadados

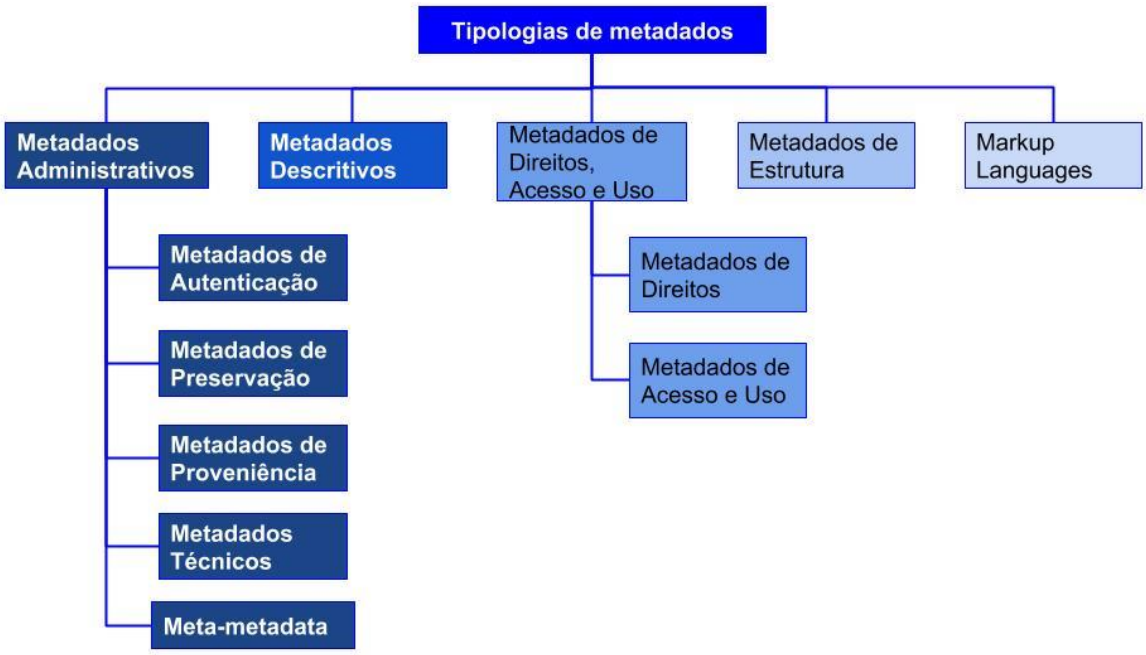
Tipo	Definição
Metadados Administrativos	Metadados administrativos são usados para gerenciar e administrar coleções e recursos informacionais, para auxiliar na tomada de decisão e manutenção dos registros e recursos informacionais. Fornecem informações sobre a origem e a manutenção de um objeto
Metadados de autenticação	Metadados de autenticação são informações que possibilitam a identificação, integridade, legitimidade de um recurso informacional. Exemplos consistem em: código de identificação ou verificação, assinatura digital, entre outros. (GREENBERG, 2001).
Metadados Preservação	Metadados de preservação estão relacionados com informações de preservação e conservação dos recursos informacionais.
Metadados de Proveniência	Metadados de proveniência estão relacionadas às informações de procedência, fornece dados sobre entidades, criação e modificações e seus relacionamentos. (POMERANTZ, 2015).
Metadados Técnicos	Metadados técnicos estão relacionados a como um sistema funciona, fornecendo informações do sistema ou do recurso.
Meta-metadata	Meta-metadata corresponde à informações sobre o registro criado, ou informações da criação de um conjunto de dados.
Metadados Descritivos	Metadados descritivos descrevem características identificadoras e os contextos intelectuais dos recursos de informação para fins de descoberta, identificação, seleção, aquisição, contexto e compreensão. (JOUDREY; TAYLOR, 2018).
Metadados de Direitos	Metadados de direitos estão relacionados às informações sobre propriedade, e direitos autorais.

Metadados de acesso e uso	Metadados acesso e uso são informações de como um recurso informacional foi acessado e usado, como restrições de circulação e acesso, registros de exposições, entre outros.
Metadados estruturais	Metadados estruturais está relacionado à composição e organização do recurso informacional.
Markup languages	Markup languages integra metadados e sinalizações para outros recursos estruturais ou semânticos. (RILEY, 2017).

Fonte: Elaborado pelo autor.

A partir da definição do quadro 11, foi possível estabelecer a estruturação da tipologia de metadados, conforme apresentado na figura 8.

Figura 8 - Tipologia de metadados



Fonte: Elaborado pelo autor.

Conforme destacado na literatura, metadados de proveniência, preservação e técnicos poderiam ser categorias a parte, inclusive possuem padrões específicos para descrição das informações, entretanto, foram mantidos como subcategorias dos metadados administrativos, pois eles são indissociáveis ao gerenciamento e à administração dos recursos informacionais.

Em termos gerais, todos os metadados são utilizados para auxiliar a gestão do sistema, entretanto, especificamente os metadados descritivos, de acesso e uso, estruturais e *Markup Languages* possuem funções primárias que podem sobrepor às questões

administrativas e gerenciais de um sistema.

5 PROVENIÊNCIA NOS PADRÕES DE METADADOS NO DOMÍNIO BIBLIOGRÁFICO

Neste capítulo será abordada a questão da proveniência e o *crosswalk* nos padrões de metadados BIBFRAME, MARC21, *Dublin Core*, *Schema.org*, e PREMIS com PROV-O. Entretanto, antes da apresentação do *crosswalk*, foi realizada uma contextualização dos ambientes digitais, a fim de identificar os possíveis ambientes de aplicação da pesquisa.

Os ambientes informacionais digitais configuram-se como sistema, sistema de informação, *sites*, *Websites*, portais, espaços de informação, ambiente de informação, ambiente digital, *softwares*, aplicações, entre outros. (CAMARGO; VIDOTTI, 2011). Nesse contexto, Bisset Alvarez (2017) destaca que com a evolução das tecnologias de informação e da *Web*, foram criados diversos sistemas que contribuíram para o processo de automação das bibliotecas. Entre eles destacam-se os sistemas de gerenciamento de bibliotecas, os repositórios digitais, bibliotecas digitais, Sistemas de Gerenciamento de Revistas Eletrônicas de gerenciamento de bibliotecas e Sistemas de busca e descoberta. Apesar dos catálogos estarem atrelados na literatura principalmente aos sistemas de gerenciamento de bibliotecas, destaca-se que cada sistema apontado por Bisset Alvarez (2017) possui catálogos com características similares.

A evolução dos catálogos e da catalogação foram destacados por diversos autores como (ALVES, Rachel Cristina Vesu; SANTOS, 2013; BALBY, 1995; BARBOSA, 1978; BASTOS, 2013; CASTRO; SANTOS, 2014; MEY, 1995; MEY; SILVEIRA, 2009; SANTOS, P.; PEREIRA, 2014); entre outros. Dessa forma, de acordo com Mey e Silveira (2009, p. 12)

> Catálogo é um meio de comunicação que veicula mensagens sobre os registros de conhecimento, de um ou vários acervos, reais ou ciberespaciais, apresentando-se com sintaxe e semântica próprias e reunindo os registros do conhecimento por semelhanças para os usuários desses acervos. O catálogo explicita, por meio das mensagens, os atributos das entidades e os relacionamentos entre elas.

Nessa perspectiva, o catálogo que possibilita o usuário localizar e identificar um item na biblioteca. "Em síntese, pode-se dizer que os catálogos veiculam as mensagens elaboradas pela catalogação, permitindo aos usuários encontrar os registros do conhecimento de seu interesse e permitindo aos registros do conhecimento encontrar seus usuários." (MEY; SILVEIRA, 2009, p. 13).

O processo de automação dos catálogos resultou em novas possibilidades de acesso aos usuários e com o surgimento dos *Online Public Access Catalog* (OPAC) tornou possível a consulta dos itens de uma biblioteca, por acesso remoto. Segundo Thanuskodi (2012, p. 70, tradução nossa), "O Catálogo de Acesso Público Online (OPAC) é um sistema de recuperação de informação caracterizado por registros bibliográficos, principalmente de livros, periódicos e materiais audiovisuais disponíveis em uma biblioteca particular."

Balby (2002) caracteriza os OPACs como interfaces de bases de dados que permitem aos usuários realizar busca e localização de recursos informacionais. Silva, E. e Boccato (2012) complementam que os OPACs são sistemas de recuperação da informação que permitem a busca, a localização e o acesso de recursos informacionais por meio de pontos de acesso como autor, título, assunto, data, local, entre outros.

> Os Catálogos de Acesso Público Online (OPACs) existentes demonstram diferenças no intervalo e complexidade de suas caracterísricas funcionais, terminologia e recursos de ajuda. Enquanto muitas bibliotecas já possuem OPACs, há a necessidade de reunir, na forma de diretrizes ou recomendações, um corpus de boas práticas para ajudar as bibliotecas a projetar ou redesenhar os displays para seus OPACs, levando em consideração as necessidades dos usuários. (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2005, p. 6, tradução nossa).

A evolução dos OPACs já foi discutida principalmente por Balby (2002) e Silva, E. e Boccato (2012). De acordo com a *International Federation of Library Associations and Institutions* (2005, p. 6, tradução nossa) a história dos OPACs foi influenciada principalmente por um sistema centralizado "[...] projetados e controlado por projetistas de sistemas e programadores, para mais e mais distribuído e sistemas personalizáveis. Ainda existem várias áreas nas quais os sistemas podem ser melhorados em termos de *displays*."

> Um OPAC contém todas as informações bibliográficas de um centro de informação [...] OPAC é a forma moderna e flexível do catálogo, geralmente com acesso instantâneo e sofisticado para qualquer informação gravada dentro de um computador. (SANKARI et al., 2013, p. 18, tradução nossa).

Entretanto, com o surgimento dos sistemas de descobertas, os catálogos começaram a integrar diversos sistemas em um único ambiente. Os sistemas de descoberta, serviços de descoberta ou ainda Plataformas de serviços de bibliotecas (*Library Services Platforms*), foram tratados por Breeding (2013, 2015), Maranhão (2012), Bastos (2013), Fonseca e Andrade (2014) e Pavão (2014). De acordo com Breending (2013) os sistemas de descoberta

surgiram para fornecer formas de acesso a diversas coleções de bibliotecas e serviços fornecidos por sistemas de bibliotecas integradas. Para Maranhão (2012, p. 02) "[...] serviços de descoberta utilizam a tecnologia de *harvesting*, coletando e reunindo em um índice único metadados e, às vezes, texto completo de diversas fontes."

> Assim, para se manter uma ferramenta relevante diante da diversidade de mecanismos de busca e multiplicidade de fontes informacionais, o catálogo, no contexto das bibliotecas, passa a oferecer a seus usuários a sensação de busca na Internet através dos recursos das tecnologias *Web* 2.0, agregando coleções multimídias por meio de uma interface de busca integrada que faz parte dos chamados sistemas de descoberta. (BASTOS, 2013, p. 45).

Conforme relatado por Maranhão (2012), Pavão (2014) e Araújo et al. (2015) há diversas empresas oferecendo esse serviço como o *EBSCO Discovery Service* (EDS) da *EBSCO Information Services*, o Primo/Primo Central do *ExLibris Group*, o Summon, desenvolvido pela *Proquest* e o *WorldCat Discovery da Online Computer Library* Center (OCLC).

> Com intuito de prover o acesso em todos aspectos às coleções das bibliotecas, ou seja, não somente aos materiais do catálogo da biblioteca tradicional, as ferramentas de descoberta passam a fazer parte dos sistemas integrados de bibliotecas, ajudando os usuários a descobrirem conteúdos disponíveis na biblioteca independente do formato – impresso ou eletrônico, e do acesso – local ou remoto para os materiais assinados mantidos pelas instituições. (BASTOS, 2013, p. 58-59)

Alguns sistemas de descoberta desenvolveram suas buscas e os critérios de relevâncias a partir da recuperação pela navegação facetada. Segundo Silva e Lima (2015, p. 3) a navegação facetada está relacionada "[...] à técnica da navegação facetada (o engenho); ao ato do usuário ao realizar a navegação facetada (ação de recuperar a informação); ou à interface ou aplicativo usado para realizar a navegação facetada (instrumento)."

Nelson e Turney (2015) fazem uma ressalva em relação à busca facetada e sua diferença com filtros. Uma vez que são frequentemente confundidos e utilizados como sinônimos.

> A diferença fundamental entre uma faceta e um filtro de intervalo é que os termos encontrados em uma faceta são indexados, enquanto um filtro de intervalo (por exemplo, a data ou preço) não é um termo indexado. É importante manter uma distinção clara entre uma busca facetada e um filtro porque os metadados subjacentes são diferentes. (NELSON; TURNEY, 2015, p. 79, tradução nossa).

Assim sendo, o usuário realiza uma busca no serviço de descoberta, o sistema busca

na base interna de dados que indexa bases de dados comerciais, o catálogo de bibliotecas e outras fontes na internet que a instituição deseja que recupera a informação. Assim, é possível explanar que o sistema de descoberta atua como uma plataforma mediadora entre o usuário, e intermedia a partir dos metadados os resultados entre as bases de dados comerciais, catálogos e fontes abertas na internet. Vaughan (2011) apresenta as principais funções de um sistema de descoberta.

> - Conteúdo. Estes conteúdos serviços de colheita de repositórios locais e hospedados remotamente e criar um índice de até centralizada muito abrangente o artigo nível baseada em um esquema normalizado em todos os tipos de conteúdo, bem adequado para pesquisa rápida e recuperação de resultados classificados por relevância. O conteúdo é ativado através da exploração dos recursos biblioteca local, combinada com acordos *brokered* com editoras e agregadores permitindo o acesso a seus metadados e / ou conteúdo de texto completo para fins de indexação.
> - Discovery. Estes serviços têm uma única caixa de busca proporcionar uma experiência de pesquisa *Google-like* (bem como capacidades de pesquisa avançadas).
> - Entrega. Esses serviços fornecem classificados por relevância em uma funcionalidade de interface oferta moderna e sugestões do projeto intuitivo e esperadas pelos usuários de hoje resultados rápidos (como a navegação facetada para detalhar resultados mais específicos).
> - Flexibilidade. Esses serviços são agnósticos aos sistemas subjacentes, sejam hospedados pela biblioteca ou hospedados remotamente por provedores de conteúdo. Estes serviços estão abertos em comparação com sistemas de bibliotecas tradicionais e permitir uma biblioteca maior latitude para personalizar os serviços e fazer o serviço própria. (VAUGHAN, 2011b)

Pavão (2014) e Vaughan (2012) apontam características essenciais de um sistema de descoberta. Para eles, um sistema de descoberta precisa:

> a) possuir uma interface de pesquisa unificada para o usuário, que reúna a informação do catálogo da biblioteca e de outras fontes como artigos de periódicos, imagens, materiais de arquivo, etc. b) possibilitar a descoberta mais ampla possível dos recursos da biblioteca; c) ser intuitivo, minimizando as habilidades, tempo e esforço necessário para os usuários descobrirem os recursos; d) permitir alto nível de personalização local e de interoperabilidade para facilitar a conexão e a troca de dados com outros sistemas que são parte da infraestrutura de informação da biblioteca, e e) demonstrar compromisso com sustentabilidade e futuras melhorias. (PAVÃO, 2014, p. 53).

Bisset Alvarez (2017) fez uma abordagem da evolução dos sistemas de recomendações, destacando seus conceitos e funcionamento. Como não é o foco do

trabalho discutir a evolução dos sistemas de recomendações, mas discutir os sistemas de recomendações como um ambiente que pode ser aplicado os conceitos de proveniência.

Outro ambiente que pode ser aplicado os conceitos de proveniência são os repositórios digitais. De acordo com Santarem Segundo (2010, p. 151) “Os repositórios digitais são sistemas de informação que facilitam a publicação e o armazenamento de documentos, além de fornecer serviços de informação, e por isso o interesse em contribuir com a organização de sua informação.” Torino (2017, p. 94) complementa que

> Os repositórios digitais (RDs) são sistemas de informação abertos e interoperáveis destinados à gestão da informação científica e acadêmica, capazes de armazenar arquivos de diversos formatos, constituindo-se em vias alternativas de comunicação científica e ampliação de visibilidade da produção.

Os repositórios digitais são definidos a partir de suas características, aplicação e objetivos aos quais se destina (TORINO, 2017, p. 95) e são caracterizados principalmente pelas publicações *e-print*, ou seja, publicação de teses, dissertações, artigos de periódicos, trabalhos em eventos, entre outros. São frequentemente divididos por repositório institucional e temáticos. Entretanto, com o surgimento e a necessidade de disponibilização de formatos abertos e recuperáveis, diversas outras categorias têm surgido, como repositório arquivístico e repositório de dados, conforme discutido por Sayão e Sales (2016) e Sanchez (2018).

Segundo Leite et al. (2009, p. 21) um repositório institucional é “[...] um serviço de informação científica - em ambiente digital e interoperável - dedicado ao gerenciamento da produção intelectual de uma instituição.” Os repositórios temáticos correspondem ao gerenciamento de produções agrupadas pela especificidade de um assunto em comum, ou melhor, área do conhecimento.

Pensando na questão da padronização e requisitos mínimos para repositórios digitais, a *International Organization for Standardization* (ISO) criou a norma 16363 de 2012 que “[...] define uma prática recomendada para avaliar a confiabilidade dos repositórios digitais. É aplicável a toda a gama de repositórios digitais. ISO 16363: 2012 pode ser usado como base para a certificação.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2012, não paginado) Ela está estruturada em três dimensões: Infraestrutura Organizacional, Gestão de Objetos Digitais, e Infraestrutura e Gestão de Segurança de Riscos. Segundo o *Research Libraries Group* e a *Online Computer Library Center* (2002, p. 5, tradução nossa)

"[...] repositório digital confiável é aquele que tem como missão oferecer, à sua comunidade-alvo, acesso confiável e de longo prazo aos recursos digitais por ele gerenciados, agora e no futuro."

Nesse contexto, segundo o Conselho Nacional de Arquivos (2015, p. 9) complementa que um repositório digital confiável "[...] é capaz de manter autênticos os materiais digitais, de preservá-los e prover acesso a eles pelo tempo necessário." De acordo com a *Research Libraries Group* e *Online Computer Library Center* (2002, p. 5) e o Conselho Nacional de Arquivos (2015, p. 9) os repositórios digitais confiáveis devem atender aos seguintes requisitos:

- aceitar, em nome de seus depositantes, a responsabilidade pela manutenção dos materiais digitais;
- dispor de uma estrutura organizacional que apoie não somente a viabilidade de longo prazo dos próprios repositórios, mas também dos materiais digitais sob sua responsabilidade;
- demonstrar sustentabilidade econômica e transparência administrativa;
- projetar seus sistemas de acordo com convenções e padrões comumente aceitos, no sentido de assegurar, de forma contínua, a gestão, o acesso e a segurança dos materiais depositados;
- estabelecer metodologias para avaliação dos sistemas que considerem as expectativas de confiabilidade esperadas pela comunidade;
- considerar, para desempenhar suas responsabilidades de longo prazo, os depositários e os usuários de forma aberta e explícita;
- dispor de políticas, práticas e desempenho que possam ser auditáveis e mensuráveis; e
- observar os seguintes fatores relativos às responsabilidades organizacionais e de curadoria dos repositórios: escopo dos materiais depositados, gerenciamento do ciclo de vida e preservação, atuação junto a uma ampla gama de parceiros, questões legais relacionadas com a propriedade dos materiais armazenados e implicações financeiras.

Segundo Thomaz (2007, p. 81), a propriedade de confiabilidade em repositórios digitais está relacionada à

- confiança de que os produtores estão enviando as informações corretas,
- confiança de que os consumidores estão recebendo as informações corretas, e
- confiança de que os fornecedores estão prestando serviços adequados.

Nesse sentido, a confiabilidade está relacionada ao documento desde a sua produção à veracidade do seu conteúdo. Isto é, a confiabilidade está atrelada a veracidade do documento, que segundo Santos e Flores (2015, p. 204) destacam que a propriedade de confiabilidade depende de como o documento é produzido e sua veracidade; e completa, baseado no Conselho Nacional de Arquivos (2011, p. 21) que "A confiabilidade é uma questão de grau, ou seja, um documento pode ser mais ou menos confiável. [...] Desta forma, entende-se, que não se pode tratar a confiabilidade como um status de 'confiável' e 'não confiável', e sim como uma variável que depende do contexto tecnológico onde está situado o acervo. (SANTOS, H.; FLORES, 2015, p.20).

Pensando nesse cenário, Carvalho et al. (2014) realizaram um estudo de auditoria em 26 repositórios institucionais de Portugal com o intuito de identificar se esses repositórios estão de acordo com a ISO 16363. Similar ao estudo de Carvalho et al. (2014) e Rezende, Cruz-Riascos e Hott (2017) apresentaram um panorama de 17 repositórios brasileiros e destacou que os repositórios analisados precisam atentar à questão da preservação digital e que nenhum repositório analisado contempla ou atende minimamente aos aspectos de infraestrutura organizacional com base na norma ISO 16363/2012. Ambos os trabalhos (CARVALHO et al., 2014; REZENDE; CRUZ-RIASCOS; HOTT, 2017) relataram uma preocupação com a questão da preservação digital dos repositórios analisados.

Entre um dos requisitos para conseguir garantir a integridade é a utilização de padrões de metadados. De acordo com Andrade e Baptista (2015) e Baptista (2017) pouco repositórios têm definido perfis de aplicação e muito utilizam apenas o Dublin Core simples. Ainda segundo Baptista (2017), "[...] os repositórios não explicitam claramente como interpretar os valores associados às propriedades em cada um dos seus registros [...]".

Nesse contexto, diversos padrões são utilizados no domínio bibliográfico. Alguns são de uso *Web* como o *Dublin Core* e outros específicos do domínio bibliográfico como MARC21 bibliográfico. Nessa seção, foram apresentados alguns padrões no domínio bibliográfico. A princípio, foi abordado o MARC21 por sua origem ser da década de 60 e, atualmente é um dos padrões mais utilizados no domínio bibliográfico. Logo em seguida, foi apresentado

Dublin Core, por ser o primeiro padrão para localização dos recursos informacionais no ambiente *Web* na década de 90.

Neste subcapítulo, será discutido a questão da proveniência nos padrões de metadados no domínio bibliográfico. Destaca-se que o intuito não é comparar os de forma a dizer qual é melhor ou pior, mas de verificar a compatibilidade e possibilidade de uso do PROV com os padrões aplicados no domínio bibliográfico.

5.1 Formato MARC21

Antes da concretização do MARC21, diversas estruturas para troca de informações estavam em vigor em todo mundo, cada uma com sua própria configuração de metadados, conforme apresentado por Santos e Pereira (2014). No ano de 1994, liderados pelos administradores do MARC dos Estados Unidos, Canadá e Inglaterra, decidiram montar uma estrutura única para facilitar o intercâmbio de dados. Após diversas tentativas, a junção do MARC dos Estados Unidos e Canadá resultou no MARC do século XXI, conhecido como MARC21. O MARC21 possui cinco formatos com usos e finalidades distintas, entretanto, o foco deste trabalho é o Formato MARC 21 para Dados Bibliográficos, daqui em diante mencionado apenas como MARC 21.

A composição do MARC é constituída por: estrutura do registro, indicação de conteúdo e conteúdo dos elementos que compõem o registro. A estrutura do registro é uma implementação do padrão internacional Formato para Intercâmbio de Informações (ISO 2709) e do protocolo de intercâmbio de informação bibliográfica (ANSI/NISO Z39.2). Os códigos e convenções estabelecidas explicitamente para identificar e caracterizar adicionalmente os elementos de dados dentro de um disco e para suportar a manipulação desses dados é definida por cada um dos formatos MARC. O conteúdo dos elementos de dados de um registro MARC é definido por padrões externos, como *International Standard Bibliographic Description* (ISBD), *Anglo-American Cataloguing Rule*s (AACR), entre outros. (LIBRARY OF CONGRESS, 2006).

A estrutura básica do MARC21 advém da ISO 2709 e é composta por: Líder, Diretório e Campos variáveis.

- Líder – “Dados que fornecem informações para o processamento do registro.” (FERREIRA, M., 2002, p. iii);

- Diretório – “Uma série de entradas que contém a posição inicial e o tamanho de cada etiqueta (TAG) dentro do registro bibliográfico.” (FERREIRA, M., 2002 p. iv);
- Campos variáveis – “Os dados em um registro bibliográfico MARC21 estão organizados em campos variáveis, cada um identificado por uma etiqueta de 3 caracteres numéricos, registrados na entrada do diretório, referente a cada campo.” (FERREIRA, M., 2002, p. iv).

O histórico do MARC já foi abordado por diversos autores como Chowdhury e Chowdhury (2007), Gonzales (2014), Santos, P. e Pereira (2014) e Assumpção e Santos (2015), entre outros.

A questão da proveniência no MARC foi pouco abordada. Quando registros MARC começaram a serem compartilhados pela *Library of Congress* na década de 60, algumas informações foram utilizadas para verificar a proveniência das informações, como número de controle do registro, entre outras informações.

Atualmente, os trabalhos localizados sobre proveniência no MARC21, discutem principalmente no contexto da custódia do recurso informacional descrito no campo 561 “História de titularidade e Custódia”.

No ano de 2012, a *Deutsche Nationalbibliothek*, *Library of Congress* (EUA), e a OCLC realizaram uma revisão e criação de alguns campos que pudessem representar a questão da proveniência em registros MARC21. Após diversas discussões foi criado o campo “883 - *Machine-generated Metadata Provenance*” em 2012, com intuito de representar algumas informações de proveniência. O campo foi definido como:

> Usado para fornecer informações sobre a proveniência de metadados em campos de dados no registro, com provisão especial para geração de máquinas. O campo 883 contém um link para o campo ao qual pertence. Destinado ao uso com campos de dados que foram total ou parcialmente gerados por máquina, ou seja, gerados por algum processo nomeado diferente da criação intelectual. (LIBRARY OF CONGRESS, 2017, não paginado, tradução nossa).

Foram definidos 10 subcampos, conforme descrito no quadro 12.

Quadro 12 - Subcampos do campo 883 MARC21, “Machine-generated Metadata Provenance”

Subcampo	Definição

$a - Generation process (NR)	“Identifica o processo usado para produzir os dados contidos no campo vinculado. O subcampo pode conter um nome de processo ou alguma outra descrição.”
$c - Confidence value (NR)	“Descreve a confiança da agência usando o processo / atividade identificado no subcampo $ a para gerar o campo vinculado. O subcampo contém um valor de ponto flutuante entre 0 e 1. Uma vírgula ou um ponto pode ser usado como um marcador decimal. 0 significa sem confiança e 1 significa confiança total.”
$d - Generation date (NR)	“Data em que o campo vinculado foi gerado. Isso também serve como o início do período de validade. A data é registrada no formato yyyymmdd de acordo com a ISO 8601, Representação de Datas e Tempos.”
$q - Generation agency (NR)	“Código de organização MARC da instituição usando o processo / atividade para gerar o campo vinculado.”
$x - Validity end date (NR)	“Data que representa o final esperado do período de validade dos dados no campo vinculado. A data é registrada no formato yyyymmdd de acordo com a ISO 8601, Representação de Datas e Tempos .”
$u - Uniform Resource Identifier (NR)	“Uniform Resource Identifier (URI), por exemplo, um URL ou URN, que identifica o processo usado para produzir os dados contidos no campo ao qual o 883 está vinculado. O URI pode levar a uma descrição textual ou estruturada do processo, ou o URL usado para gerar o conteúdo do campo vinculado pode ser fornecido diretamente, ou seja, um URL invocando um serviço da *Web* ou transmitindo uma chamada de API.”
$w - Bibliographic record control number	“Número de controle do registro bibliográfico do qual os dados no campo vinculado foram obtidos.”
$0 - Authority record control number or standard number	“Número de controle do registro de autoridade do qual os dados no campo vinculado foram obtidos”
$1 - Real World Object URI (R)	URI do mundo real
$8 - Field link and sequence number (R)	Link de campo e número de sequência

Fonte: Baseado em Library of Congress (2017)

Diante da pouca discussão de informações necessárias para proveniência no MARC21, foi realizado um mapeamento dos campos e subcampos do MARC21 com a ontologia PROV-O, para verificar a compatibilidade do MARC21 e PROV-O, conforme apresentado no quadro 13 .

Quadro 13 - *Crosswalk* PROV-O para MARC21

PROV	MARC	COMENTÁRIOS
prov:Entity	-	No MARC21 não há um campo específico para essa classe
prov:Activity	583 - Action Note (R)	A classe Atividade do PROV pode ter correspondência com o campos 583 - Action Note, pois, ambos possuem o propósito de registrar ações de processamento, referência e preservação, entre outras informações.

prov:Agent	100 - Main Entry - Personal Name (NR)	A classe agente do PROV foi mapeada para campos que refletissem a questão da indicação de autoridade de pessoa ou entidade coletiva.
	110 - Main Entry - Corporate Name (NR)	
	040 - Cataloging Source (NR)	
	700 - Added Entry - Personal Name (R)	
	710 - Added Entry - Corporate Name	
Classes		
prov:Collection	-	No MARC21 não há um campo específico para essa classe
prov:EmptyCollection	-	No MARC21 não há um campo específico para essa classe
prov:Bundle	-	No MARC21 não há um campo específico para essa classe
prov:Person	100 - Main Entry - Personal Name (NR)	A classe Person e os campos 100 e 700 possuem o mesmo contexto.
	700 - Added Entry - Personal Name (R)	
prov:SoftwareAgent	-	No MARC21 não há um campo específico para essa classe
prov:Organization	110 - Main Entry - Corporate Name (NR)	A classe Organization e os campos 110 e 710 possuem o mesmo contexto.
	040 - Cataloging Source (NR)	
	710 - Added Entry - Corporate Name	
prov:Location	008 - Fixed-Length Data Elements-General Information (NR) - 15-17 - Place of publication, production, or execution	A classe location no PROV, apresenta uma definição ampla, assim pode ser mapeada para diversos campos do MARC21.
	033 - Date/Time and Place of an Event (R)	
	260 - Publication, Distribution, etc. (Imprint) (R)	
	264 - Production, Publication, Distribution, Manufacture, and Copyright Notice (R)	
	370 - Associated Place (R)	
	518 - Date/Time and Place of an Event Note (R)	
	535 - Location of Originals/Duplicates Note (R)	
	544 - Location of Other Archival Materials Note (R)	
	852 - Location (R)	
	856 - Electronic Location and Access (R)	
Classe		Descrição
prov:Influence	-	No MARC21 não há um campo específico para essa classe
prov:EntityInfluence	-	No MARC21 não há um campo específico para essa classe
prov:Usage	-	No MARC21 não há um campo específico para essa classe
prov:Start	388 - Time Period of Creation (R)	A classe Star no PROV, apresenta uma

	033 - Date/Time and Place of an Event (R)	definição de início de algo, início de uma data, assim pode ser mapeada para diversos campos do MARC21.
	045 - Time Period of Content (	
	046 - Special Coded Dates	
	362 - Dates of Publication and/or Sequential Designation (R)	
	363 - Normalized Date and Sequential Designation (	
	518 - Date/Time and Place of an Event Note (R)	
prov:End	005 - Date and Time of Latest Transaction	A classe End no PROV, apresenta uma definição de fim de algo, início de uma fim, assim pode ser mapeada para diversos campos do MARC21.
	033 - Date/Time and Place of an Event (R)	
	045 - Time Period of Content (NR)	
	046 - Special Coded Dates (R)	
	362 - Dates of Publication and/or Sequential Designation (R)	
	518 - Date/Time and Place of an Event Note (R)	
prov:Derivation	251 - Version Information (R)	A classe Derivation do PROV possui uma definição ampla, assim pode ser mapeada para diversos campos do MARC21.
	380 - Form of Work (R)	
	533 - Reproduction Note (R)	
prov:PrimarySource	534 - Original Version Note (R)	A classe PrimarySource e o campo 534 possuem o mesmo contexto.
prov:Quotation	510 - Citation/References Note (R)	A classe quotation e o campo 510 possuem o mesmo contexto.
prov:Revision	-	No MARC21 não há um campo específico para essa classe
prov:ActivityInfluence	-	No MARC21 não há um campo específico para essa classe
prov:Generation	-	No MARC21 não há um campo específico para essa classe
prov:Communication	-	No MARC21 não há um campo específico para essa classe
prov:Invalidation	-	No MARC21 não há um campo específico para essa classe
prov:AgentInfluence	-	No MARC21 não há um campo específico para essa classe
prov:Attribution	-	No MARC21 não há um campo específico para essa classe
prov:Association	-	No MARC21 não há um campo específico para essa classe
prov:Plan	-	No MARC21 não há um campo específico para essa classe
prov:Delegation	-	No MARC21 não há um campo específico para essa classe
prov:InstantaneousEvent	-	No MARC21 não há um campo específico para essa classe
prov:Role	-	No MARC21 não há um campo específico para essa classe
Propriedade		Definição

prov:wasGeneratedBy	883 - Machine-generated Metadata Provenance (R) - $a - Generation process (NR)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov:wasDerivedFrom	380 - Form of Work (R) - $a - Form of work (R)	Os conceitos apresentados pela propriedade e os subcampos do MARC21 são similares.
	533 - Reproduction Note (R) - $a - Type of reproduction (NR)	
prov:wasAttributedTo	245 - Title Statement (NR) - $c - Statement of responsibility, etc. (NR)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov:startedAtTime	388 - Time Period of Creation (R) - $a - Time period of creation term (R)	Os conceitos apresentados pela propriedade e os subcampos do MARC21 são similares.
	046 - Special Coded Dates (R) - $ k - Início ou data única criada (NR	
	362 - Dates of Publication and/or Sequential Designation (R) - $a - Dates of publication and/or sequential designation (NR)	
	518 - Date/Time and Place of an Event Note (R) - $ a - Data / hora e local de uma nota de evento (NR)	
	518 - Date/Time and Place of an Event Note (R) - $ d - Data do evento (R)	
	883 - Machine-generated Metadata Provenance (R) - $d - Generation date (NR)	
prov:used	-	No MARC21 não há um campo específico para essa classe
prov:wasInformedBy	-	No MARC21 não há um campo específico para essa classe
prov:endedAtTime	046 - Special Coded Dates (R) - $ l - Data final criada (NR)	Os conceitos apresentados pela propriedade e os subcampos do MARC21 são similares.
	883 - Machine-generated Metadata Provenance (R)- $x - Validity end date (NR)	
prov:wasAssociatedWith	100 - Main Entry-Personal Name (NR) - $j - Attribution qualifier (R)	Os conceitos apresentados pela propriedade e os subcampos do MARC21 são similares.
	700 - Added Entry-Personal Name $j - Attribution qualifier (R)	
prov:actedOnBehalfOf	533 - Reproduction Note (R) - $c - Agency responsible for reproduction (R)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
Propriedade		**Descrição**
prov: alternateOf	-	No MARC21 não há um campo específico para essa propriedade
prov: specializationOf	-	No MARC21 não há um campo específico para essa propriedade
prov: generatedAtTime	-	No MARC21 não há um campo específico para essa propriedade
prov: hadPrimarySource	534 - Original Version Note (R) - $a - Type of reproduction (NR)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.

prov: value	-	No MARC21 não há um campo específico para essa propriedade
prov: wasQuotedFrom	-	No MARC21 não há um campo específico para essa propriedade
prov: wasRevisionOf	251 - Version Information (R) - $a - Version (R)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov: invalidatedAtTime	-	No MARC21 não há um campo específico para essa propriedade
prov: wasInvalidatedBy	-	No MARC21 não há um campo específico para essa propriedade
prov: hadMember	583 - Action Note (R) - $k - Action agent (R)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov: wasStartedBy	583 - Action Note (R) - $c - Time/date of action (R)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov: wasEndedBy	-	No MARC21 não há um campo específico para essa propriedade
prov:invalidated	-	No MARC21 não há um campo específico para essa propriedade
prov:influenced	-	No MARC21 não há um campo específico para essa propriedade
prov:atLocation	260 - Publication, Distribution, etc. (Imprint) (R) - $a - Place of publication, distribution, etc. (R)	Um local pode ser um local geográfico identificável (ISO 19112), mas também pode ser um local não geográfico, como um diretório, linha ou coluna. Como tal, existem várias maneiras pelas quais a localização pode ser expressa, como por uma coordenada, endereço, marco e assim por diante.
prov:generated		No MARC21 não há um campo específico para essa propriedade
Propriedade		
prov:wasInfluencedBy		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedInfluence		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedGeneration		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedDerivation	251 - Version Information (R) - $a - Version (R) 533 - Reproduction Note (R - $a - Type of reproduction (NR)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov:qualifiedPrimarySource		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedQuotation		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedRevision		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedAttribution		No MARC21 não há um campo específico para essa propriedade

prov:qualifiedInvalidation		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedStart		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedUsage		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedCommunication		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedAssociation		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedEnd		No MARC21 não há um campo específico para essa propriedade
prov:qualifiedDelegation		No MARC21 não há um campo específico para essa propriedade
prov:influencer		No MARC21 não há um campo específico para essa propriedade
prov:entity		No MARC21 não há um campo específico para essa propriedade
prov:hadUsage		No MARC21 não há um campo específico para essa propriedade
prov:hadGeneration		No MARC21 não há um campo específico para essa propriedade
prov:activity	583 - Action Note (R) - $a - Action (NR)	Os conceitos apresentados pela propriedade e o subcampo do MARC21 são similares.
prov:agent		No MARC21 não há um campo específico para essa propriedade
prov:hadPlan		No MARC21 não há um campo específico para essa propriedade
prov:hadActivity		No MARC21 não há um campo específico para essa propriedade
prov:atTime		No MARC21 não há um campo específico para essa propriedade
prov:hadRole		No MARC21 não há um campo específico para essa propriedade

Fonte: Elaborado pelo autor.

Os campos foram mapeados para classes do PROV pois entende-se que eles possuem funções semelhantes, o mesmo ocorre com os subcampos do MARC21, que foram mapeados com as propriedades.

O levantamento revelou que foi possível identificar correspondência de 10 classes/subclasses definidas no PROV-O, com o MARC21 e 20 classes/subclasses não tiveram

correspondência com nenhum dos campos do MARC21. Em relação às propriedades/subpropriedades, 23 propriedades/subpropriedades tiveram alguma identificação com subcampos do MARC21 e 34 propriedades não tiveram nenhuma correspondência com subcampos do MARC21.

Apesar do MARC21 ter pelo menos 1 campo específico para questões de proveniência (Campos 883) nem todas informações do campo foram possíveis de serem mapeadas para PROV e que poderiam ampliar as propriedades do PROV.

5.2 Dublin Core

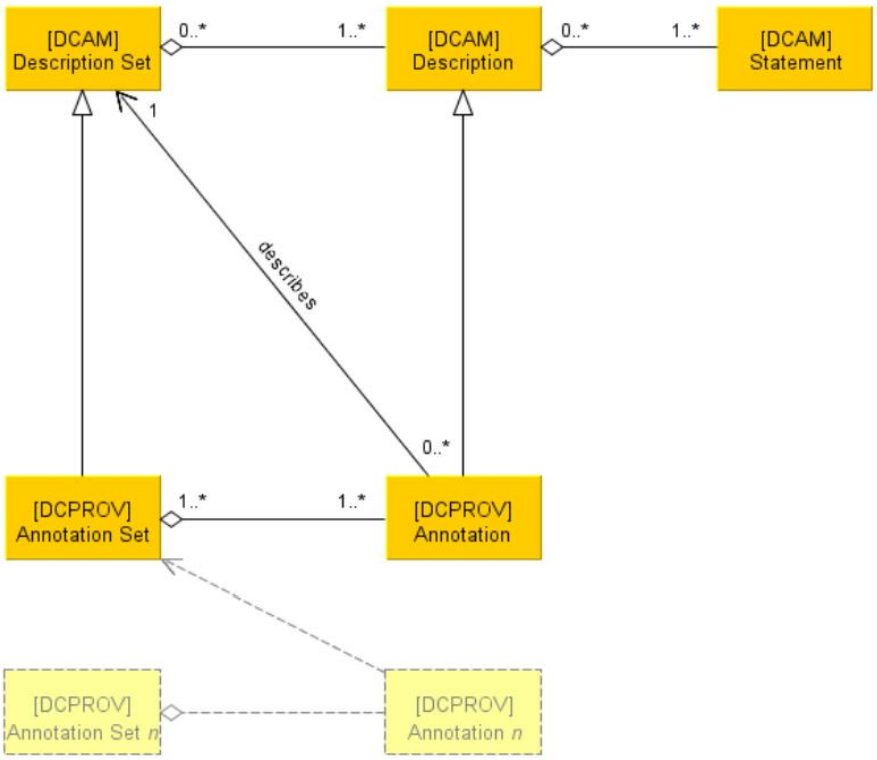
O *Dublin Core* surgiu a partir da necessidade de localizar objetos digitais na *Web* em um evento no ano de 1995 e seu histórico foi relatado por autores como Baker (2012) e Arakaki, Alves e Santos, P. (2018).

Ao longo dos anos, o *Dublin Core* teve diversas contribuições como a definição dos 15 elementos básicos para descrever qualquer objeto digital. O desenvolvimento da gramática do *Dublin Core* proposta por Baker (2000) que posteriormente estruturou os elementos qualificados do *Dublin Core*. Posteriormente, houve a proposta de padronização dos sistemas de informação com o *Dublin Core Abstract Model* (DCAM - Modelo Abstrato Dublin Core), por Powell et al. (2007) e com Nilsson, Baker e Johnston (2008) pela estruturação de perfis de aplicação, *Dublin Core Application Profiles* (DCAP - Perfil de Aplicação Dublin Core).

De acordo com Eckert, Garijo e Panzer (2011), logo na década de 90 foram realizadas algumas discussões para definir um vocabulário e diretrizes de metadados de proveniência denominado de ACore que posteriormente virou o *Administrative Components* (AC) que abordavam metadados para todo o registro, para atualização e alteração, e para intercâmbio de registros em lote (HANSEN; ANDRESEN, 2003), entretanto, não havia um modelo para relacionar as informações de proveniência com os metadados.

Como processo de ampliação e compatibilidade de aplicação do Dublin Core com a questão da proveniência, Eckert, Garijo e Panzer (2011) apresentaram uma arquitetura para incluir informações de proveniência incorporadas ao DCAM. A proposta consiste em relacionar a estrutura abstrata das informações de proveniência aos elementos das entidades existentes do DCAM, conforme apresentado na figura 9.

Figura 9 - DCAM e a proveniência



Fonte: Eckert, Garijo e Panzer (2011, p. 15)

O modelo ampliado do DCAM utiliza os seguintes elementos:

> Descrição Set (da terminologia DCAM11): Um conjunto de uma ou mais Descrições, cada uma dos quais descreve um único recurso.
> Descrição (da terminologia DCAM): uma ou mais declarações sobre uma e somente um recurso.
> Statement12 (da terminologia DCAM): uma instanciação de um par de valor de propriedade feita up de um URI de propriedade (um URI que identifica uma propriedade) e um substituto de valor.
> Anotação: Uma ou mais declarações sobre um conjunto de descrição. Subclasse de Descrição.
> Conjunto de anotações: um conjunto de uma ou mais anotações. Subclasse do Conjunto de Descrição.[4] (ECKERT; GARIJO; PANZER, 2011, p. 14, tradução nossa).

Para comprovação do modelo, Eckert, Garijo e Pazner (2011, p. 14, tradução nossa) fizeram testes utilizando o protocolo *Open Archives Initiative Protocol for Metadata*

4 Citação do original: "*Description Set (from DCAM terminology11): A set of one or more Descriptions, each of which describes a single resource. Description (from DCAM terminology): One or more Statements about one, and only one, resource. Statement12 (from DCAM terminology): An instantiation of a property-value pair made up of a property URI (a URI that identifies a property) and a value surrogate. Annotation: One or more Statements about one Description Set. Subclass of Description. Annotation Set: A set of one or more Annotations. Subclass of Description Set.abordar a web semântica e as boas praticas de publicação de dados abertos e linked data.*" (ECKERT; GARIJO; PANZER, 2011, p. 14).

Harvesting (OAI-PMH) para estrutura proposta em DC-PROV, utilizando os seguintes elementos

> - baseURL: o baseURL do repositório de origem do qual o registro de metadados foi colhido
> - identificador: o identificador único do item no repositório de origem a partir do qual o registro de metadados foi divulgado
> - datestamp: o registro de data e data do registro de metadados disseminado pela origem repositório
> - metadataNamespace: o URI de namespace XML do formato de metadados do registro colhida do repositório de origem
> - originDescription: um bloco originDescription opcional que foi obtido quando o registro de metadados foi coletado. Um conjunto de blocos originDescription aninhados descrever a procedência sobre uma sequência de colheitas
> - harvestDate: o responseDate da resposta OAI-PMH que resultou no registro
> - sendo colhida do repositório de origem
> - alterado: um valor booleano que deve ser verdadeiro se o registro colhido foi alterado antes sendo divulgado novamente (ECKERT; GARIJO; PANZER, 2011, p. 14, tradução nossa).

Entretanto, os autores concluíram que os metadados extraídos pelo OAI-PMH fornecem apenas a proveniência para agregações da coleta de metadados e mesmo o OAI-ORE não foi especificamente projetado para representar informações de proveniência. (ECKERT; GARIJO; PANZER, 2011, p. 13, tradução nossa).

Fruto dos trabalhos de Eckert, Garijo e Panzer, dentre outros, Garijo e Eckert (2013) apresentaram com documento da família PROV, um mapeamento das classes/subclasses, propriedades/subpropriedades do PROV para o DC *terms*, conforme apresentado no quadro 14.

Quadro 14 - *Crosswalk* PROV-O para DCTerm

PROV Term	DC Term	Relação
prov:Agent	**dct:Agent**	Ambos dct: Agent e prov: Agent referem-se ao mesmo conceito.
prov:Location	**dct:Location**	Ambos dct: Location e prov: a localização referem-se ao mesmo conceito.
prov:Bundle	**dct:ProvenanceStatement**	A dct: ProvenanceStatement é definido como "Uma declaração de quaisquer alterações na propriedade e custódia de um recurso desde a sua criação", que é um contêiner para qualquer declaração relacionada à proveniência.

prov:Entity	**dct:BibliographicResource**	Um recurso bibliográfico refere-se a livros, artigos, etc., que são entidades PROV concretas.
prov:Entity	**dct:LicenseDocument**	Documento concedendo permissão para fazer algo em relação a um recurso. Assim, é mapeado como um tipo de prov: Entidade.
prov:Entity	**dct:RightsStatement**	Declaração sobre os direitos intelectuais de um recurso (por exemplo, um documento). Assim, é mapeado como um tipo de prov: Entidade.
prov:Entity	**dct:PhysicalResource**	Uma coisa material, que é um tipo concreto de prov: Entidade.
prov:Location	**dct:LocationPeriodOrJurisdiction**	dct: LocationPeriodOrJurisdiction é uma superclasse de dct: Location (equivalente a prov: Location).
prov:Plan	**dct:LinguisticSystem**	A dct: LinguisticSystem é um sistema de símbolos, sons, gestos, etc. usado na comunicação. Portanto, o sistema linguístico define o plano a seguir para aprender uma determinada língua.
prov:Plan	**dct:MethodOfAccrual**	dct: MethodOfAccrual define o método pelo qual os itens são adicionados a uma coleção (ou seja, o prov: Plano seguido na atividade de inserção).
prov:Plan	**dct:MethodOfInstruction**	Processo que é usado para gerar conhecimento, atitude e habilidades. Como dct: MethodOfInstruction define o método associado a uma atividade, ele é mapeado como subclasse de prov: Plan.
prov:Plan	**dct:Policy**	dct: Política é definida como "um plano ou curso de ação por uma autoridade, destinado a influenciar e determinar decisões, ações e outros assuntos". Esta é uma especialização de prov: Plan.
prov:alternateOf	**dct:hasFormat**	Veja a justificativa para dct: isFormatOf (como prov: alternateOf).
prov:alternateOf	**dct:isFormatOf**	dct: isFormatOf refere-se a outro recurso que é o mesmo, mas em outro formato. Assim, o termo é mapeado para prov: alternateOf.
prov:generatedAtTime	**dct:created**	Propriedade usada para descrever o tempo de criação de um recurso (ou seja, o tempo de sua geração). Nós o mapeamos como uma subpropriedade de prov: generatedAtTime porque "criação" é uma das muitas atividades que geram uma entidade (por exemplo, geração inclui modificação, emissão, aceitação, etc.).
prov:generatedAtTime	**dct:dateAccepted**	Propriedade usada para descrever a data em que o recurso foi aceito. dct: dateAccepted é mapeado como uma subproperty de prov: generatedAtTime porque o recurso aceito foi gerado por uma atividade "Accept" que pode ter

		mudado de seu estado anterior.
prov:generatedAtTime	**dct:dateCopyrighted**	Propriedade usada para descrever a data em que o recurso foi protegido por direitos autorais. dct: dateCopyrighted é mapeado como uma subproperty de prov: generatedAtTime porque o recurso protegido por direitos autorais foi gerado por uma atividade "CopyRight" que pode ter mudado de seu estado anterior.
prov:generatedAtTime	**dct:dateSubmitted**	Propriedade usada para descrever a data em que o recurso foi enviado. dct: dateSubmitted é mapeado como uma subproperty de prov: generatedAtTime porque o recurso enviado foi gerado por uma atividade "Submit" que pode ter mudado de seu estado anterior.
prov:generatedAtTime	**dct:issued**	Propriedade usada para descrever a data em que o recurso foi emitido. dct: emitido é mapeado como uma subproperty de prov: generatedAtTime porque o recurso emitido é uma entidade em si, que foi gerada em um determinado momento.
prov:generatedAtTime	**dct:modified**	Propriedade usada para descrever a data em que o recurso foi modificado. dct: modified é mapeado como uma subproperty de prov: generatedAtTime porque o recurso modificado foi gerado por uma atividade "Modify" que o alterou de seu estado anterior.
prov:hadDerivation	**dct:hasFormat**	Propriedade inversa de dct: isFormatOf.
prov:hadDerivation	**dct:isReferencedBy**	Propriedade inversa de dct: references.
prov:hadPrimarySource	**dct:source**	A definição de prov: hadPrimarySource ("algo produzido por algum agente com experiência direta e conhecimento sobre o tópico") é mais restritiva que dct: source ("Um recurso relacionado do qual o recurso descrito é derivado").
prov:hadRevision	**dct:hasVersion**	Propriedade inversa de dct: isVersionOf.
prov:has_provenance	**dct:provenance**	Propriedade usada para vincular um recurso a seu dct correspondente: ProvenanceStatement. Como prov: has_provenance pode referir-se a qualquer tipo de registro de proveniência, dct: provenience é mapeado como uma subclasse.
prov:wasAttributedTo	**dct:creator**	Um criador é um dos agentes que participaram da criação de um recurso. Eles têm a atribuição para o resultado dessa atividade.
prov:wasAttributedTo	**dct:contributor**	Um colaborador está associado à atividade de criação ou à atualização do recurso. Portanto, ele / ela tem atribuição sobre o resultado dessas atividades.
prov:wasAttributedTo	**dct:publisher**	Um editor tem a atribuição do recurso publicado

		depois de participar da atividade de publicação que o gerou.
prov:wasAttributedTo	**dct:rightsHolder**	O detentor dos direitos tem a atribuição da licença associada a um recurso. Assim, podemos dizer que o recurso é atribuído em parte ao detentor dos direitos.
prov:wasDerivedFrom	**dct:isFormatOf**	dct: isFormatOf refere-se a outro recurso "pré-existente" que é o mesmo, mas em outro formato (de acordo com o cdt: hasFormat), implicando que o novo recurso é baseado no primeiro.
prov:wasDerivedFrom	**dct:references**	No PROV, uma derivação é definida como "uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova, ou a construção de uma nova entidade baseada em uma entidade pré-existente". Se um recurso n1 fizer referência a outro recurso o1, então a construção de n1 é baseada em o1, mesmo se o1 não influenciar n1 significativamente. Remover a referência para o1 em n1 levaria à construção de outro recurso n1 ', diferente de n1.
prov:wasDerivedFrom	**dct:source**	dct: source é definido como "um recurso relacionado do qual o recurso descrito é derivado", que corresponde à noção de derivação em PROV-DM ("uma transformação de uma entidade em outra"). No entanto, prov: wasDerivedFrom também abrange derivações mais amplas, como "uma atualização de uma entidade resultando em uma nova", que não é coberta por dct: source.
prov:wasRevisionOf	**dct:isVersionOf**	Semelhante à propriedade anterior, prov: wasRevisionOf é mais restritivo no sentido de que se refere a uma versão revisada de um recurso, enquanto dct: isVersionOf envolve versões, edições ou adaptações do recurso original. Como exemplo, "West Side Story" é uma versão (adaptação) de "Romeu e Julieta", mas não uma revisão.

Fonte: Garijo e Eckert (2013, não paginado, tradução nossa)

Não foi possível realizar a correspondência de equivalência em diversas classes, subclasses, propriedades e subpropriedades entre o PROV e *Dublin Core*. Assim, Garijo e Eckert (2013) fizeram adaptações para conciliar os dois vocabulários. Nesse contexto, algumas classes, subclasses, propriedades e subpropriedades do *Dublin Core* tornaram subclasses ou subpropriedades do PROV, validando assim a compatibilidade de alguns

elementos entre os dois vocabulários.

5.3 BIBFRAME: *Bibliographic Framework Initiative*

No ano de 2011, a *Library of Congress* dos Estados Unidos iniciou a construção de um novo modelo de dados para o domínio bibliográfico com a proposta de criar um padrão substituto do MARC21 denominado *Bibliographic Framework Initiative* (BIBFRAME). De acordo com a *Library of Congress* (2012) o novo modelo apresenta características da proposta pelo *Functional Requirements for Bibliographic Records* (FRBR) e ainda os princípios do *Linked Data*. De acordo com Silva, R. (2013), o BIBFRAME é flexível, possui uma arquitetura para expressar e conectar informações, pode ser adotado para comunidades além das bibliotecas, entre outras. (ARAKAKI, 2016)

A primeira versão do BIBFRAME foi composta por um modelo que possui, basicamente, duas classes principais, Obra (*Work*) e Instância (*Instance*). A Obra é um recurso que reflete a essência conceitual de um recurso catalogado, ou seja, corresponde às entidades Obra e Expressão do FRBR. A Instância é um recurso que reflete uma forma individual de realização do material de Obra, ou seja, é a materialização da Obra do BIBFRAME. A classe Instância corresponde às entidades Manifestação e Item do FRBR. (ARAKAKI, 2016; ARAKAKI et al., 2017).

Após diversos estudos e testes, a *Library of Congress* (EUA) publicou o Modelo BIBFRAME 2.0 em abril de 2016. O Modelo BIBFRAME 2.0 consiste de três classes maiores: *Work* (Obra), *Instance* (Instância) e *Item* (Item). A Obra foi definida como o nível mais alto de abstração e manteve seu relacionamento com a Obra e Expressão do FRBR. A Instância foi caracterizada por possuir uma ou mais formas de realização de uma Obra. Uma Instância reflete informações, como seu editor, seu local e sua data de publicação e seu formato e foi caracterizada pela entidade Manifestação do FRBR. Um item é definido como uma cópia real (física ou eletrônica) de uma instância. Ele possui informações, tais como a sua localização (física ou virtual), marca de prateleira e código de barras e foi caracterizada pela entidade Item do FRBR. (ARAKAKI, 2016; ARAKAKI et al., 2017; LIBRARY OF CONGRESS, 2016).

O Modelo BIBFRAME 2.0 aborda, ainda, algumas outras classes, como *Agents* (Agentes), *Subjects* (Assuntos) e *Events* (Eventos). A classe Agente pode ser definida como pessoas, organizações, jurisdições e estão associadas a uma Obra ou Instância. A classe Agente pode ter diversas funções como: autor, editor, artista, fotógrafo, compositor, ilustrador etc. (LIBRARY OF CONGRESS, 2016). A classe Assunto foi caracterizada pelas

informações de conteúdo, ou seja, “sobre o quê” de uma Obra e pode ter um ou mais conceitos. Estes incluem temas, lugares, expressões temporais, eventos, obras, instâncias, itens, agentes, etc. (LIBRARY OF CONGRESS, 2016). A classe Evento foi definida como ocorrências que podem ser estar relacionadas ao conteúdo de uma Obra (LIBRARY OF CONGRESS, 2016, ARAKAKI, 2016).

Para completar o Modelo BIBFRAME, a *Library of Congress* disponibilizou também o vocabulário BIBFRAME com classes e propriedades de descrição e de relacionamentos que são compatíveis com o RDF. (ARAKAKI, 2016).

Pouco se tem discutido a questão da proveniência no BIBFRAME e foram identificados apenas indicações de trabalhos futuros como de Kovari, Folsom e Younes (2017). Nesse contexto, o quadro 15 apresenta o *crosswalk* do PROV-O para o BIBFRAME.

Quadro 15 - *Crosswalk* PROV-O para BIBFRAME

PROV	**BIBFRAME**	
prov:Entity	-	No BIBFRAME não foi localizada classe compatível
prov:Activity	-	No BIBFRAME não foi localizada classe compatível
prov:Agent	Agent	Os conceitos apresentados pelas classes são similares.
prov:Collection	Collection	Os conceitos apresentados pelas classes são similares.
prov:EmptyCollection	-	No BIBFRAME não foi localizada classe compatível
prov:Bundle	-	No BIBFRAME não foi localizada classe compatível
prov:Person	Person	Os conceitos apresentados pelas classes são similares.
prov:SoftwareAgent	-	No BIBFRAME não foi localizada classe compatível
prov:Organization	Organization	Os conceitos apresentados pelas classes são similares.
prov:Location	Local	Os conceitos apresentados pelas classes são similares.
Classe		No BIBFRAME não foi localizada classe compatível
prov:Influence		No BIBFRAME não foi localizada classe compatível
prov:EntityInfluence		No BIBFRAME não foi localizada classe compatível
prov:Usage	UsageAndAccessPolicy	Os conceitos apresentados pelas classes são similares.
	AccessPolicy UsePolicy RetentionPolicy	
prov:Start	creationDate	Os conceitos apresentados pelas classes são similares.
prov:End		No BIBFRAME não foi localizada classe compatível
prov:Derivation		No BIBFRAME não foi localizada classe compatível

prov:PrimarySource	Source	Os conceitos apresentados pelas classes são similares.
prov:Quotation		No BIBFRAME não foi localizada classe compatível
prov:Revision		No BIBFRAME não foi localizada classe compatível
prov:ActivityInfluence		No BIBFRAME não foi localizada classe compatível
prov:Generation	Generation GenerationProcess	Os conceitos apresentados pelas classes são similares.
prov:Communication		No BIBFRAME não foi localizada classe compatível
prov:Invalidation		No BIBFRAME não foi localizada classe compatível
prov:AgentInfluence		No BIBFRAME não foi localizada classe compatível
prov:Attribution		No BIBFRAME não foi localizada classe compatível
prov:Association		No BIBFRAME não foi localizada classe compatível
prov:Plan		No BIBFRAME não foi localizada classe compatível
prov:Delegation		No BIBFRAME não foi localizada classe compatível
prov:InstantaneousEvent		No BIBFRAME não foi localizada classe compatível
prov:Role	Role	Os conceitos apresentados pelas classes são similares.
prov:wasGeneratedBy	Generate	Os conceitos apresentados pelas propriedades são similares.
prov:wasDerivedFrom	derivativeOf ou derivedFrom	Os conceitos apresentados pelas propriedades são similares.
prov:wasAttributedTo	responsibilityStatement	Os conceitos apresentados pelas propriedades são similares.
prov:startedAtTime	creationDate	Os conceitos apresentados pelas propriedades são similares.
prov:used		No BIBFRAME não foi localizada classe compatível
prov:wasInformedBy	-	No BIBFRAME não foi localizada classe compatível
prov:endedAtTime	-	No BIBFRAME não foi localizada classe compatível
prov:wasAssociatedWith	responsibilityStatement	Os conceitos apresentados pelas propriedades são similares.
prov:actedOnBehalfOf	-	No BIBFRAME não foi localizada classe compatível
prov: alternateOf		No BIBFRAME não foi localizada classe compatível
prov: specializationOf		No BIBFRAME não foi localizada classe compatível
prov: generatedAtTime		No BIBFRAME não foi localizada classe compatível
prov: hadPrimarySource	Source	Os conceitos apresentados pelas propriedades são similares.
prov: value		No BIBFRAME não foi localizada classe compatível
prov: wasQuotedFrom	references - referencedBy	Os conceitos apresentados pelas propriedades são similares.
prov: wasRevisionOf		No BIBFRAME não foi localizada classe compatível
prov: invalidatedAtTime		No BIBFRAME não foi localizada classe compatível
prov: wasInvalidatedBy		No BIBFRAME não foi localizada classe compatível
prov: hadMember		No BIBFRAME não foi localizada classe compatível

prov: wasStartedBy		No BIBFRAME não foi localizada classe compatível
prov: wasEndedBy		No BIBFRAME não foi localizada classe compatível
prov:invalidated		No BIBFRAME não foi localizada classe compatível
prov:influenced		No BIBFRAME não foi localizada classe compatível
prov:atLocation	physicalLocation - place - sublocation	Os conceitos apresentados pelas propriedades são similares.
prov:generated		No BIBFRAME não foi localizada classe compatível
prov:wasInfluencedBy		No BIBFRAME não foi localizada classe compatível
prov:qualifiedInfluence		No BIBFRAME não foi localizada classe compatível
prov:qualifiedGeneration		No BIBFRAME não foi localizada classe compatível
prov:qualifiedDerivation		No BIBFRAME não foi localizada classe compatível
prov:qualifiedPrimarySource		No BIBFRAME não foi localizada classe compatível
prov:qualifiedQuotation		No BIBFRAME não foi localizada classe compatível
prov:qualifiedRevision		No BIBFRAME não foi localizada classe compatível
prov:qualifiedAttribution		No BIBFRAME não foi localizada classe compatível
prov:qualifiedInvalidation		No BIBFRAME não foi localizada classe compatível
prov:qualifiedStart		No BIBFRAME não foi localizada classe compatível
prov:qualifiedUsage		No BIBFRAME não foi localizada classe compatível
prov:qualifiedCommunication		No BIBFRAME não foi localizada classe compatível
prov:qualifiedAssociation		No BIBFRAME não foi localizada classe compatível
prov:qualifiedEnd		No BIBFRAME não foi localizada classe compatível
prov:qualifiedDelegation		No BIBFRAME não foi localizada classe compatível
prov:influencer		No BIBFRAME não foi localizada classe compatível
prov:entity		No BIBFRAME não foi localizada classe compatível
prov:hadUsage		No BIBFRAME não foi localizada classe compatível
prov:hadGeneration		No BIBFRAME não foi localizada classe compatível
prov:activity		No BIBFRAME não foi localizada classe compatível
prov:agent		No BIBFRAME não foi localizada classe compatível
prov:hadPlan		No BIBFRAME não foi localizada classe compatível
prov:hadActivity		No BIBFRAME não foi localizada classe compatível
prov:atTime		No BIBFRAME não foi localizada classe compatível
prov:hadRole		No BIBFRAME não foi localizada classe compatível

Fonte: Elaborado pelo autor

O mapeamento do PROV para BIBFRAME teve algumas facilidades, pois, ambas possuem a mesma terminologia de classe e propriedades. Isso facilita alguns procedimentos para o mapeamento, pois não há necessidade de adaptações entre conceitos de classes e propriedades.

O mapeamento revelou que 10 classes do PROV foram mapeadas para o BIBFRAME,

e oito (8) propriedades foram mapeadas para o BIBFRAME, sendo que algumas classes tiveram o tipo de mapeamento de um-para-muitos, e algumas propriedades tiveram seu mapeamento como muitos-para-um.

5.4 Schema.org

O *Schema.org* é uma iniciativa comunitária e colaborativa com a missão de criar, manter e promover esquemas de dados estruturados para Internet. (SCHEMA.ORG, 2011?). O *Schema.org* surgiu de uma iniciativa de buscadores, como Google, Yahoo, Bing e Yandex para desenvolver uma estrutura que seja capaz de melhorar a busca de informações.

Segundo Pomerantz (2015) o *Schema.org* é baseado em microdados, que é uma especificação para a incorporação de metadados dentro de uma página da *Web*. Atualmente o *Schema.org* está na versão 3.5 e pode ser utilizado com diversas codificações como RDFa, Microdata e JSON-LD.

De acordo com Roa-Martínez, Vidotti e Pastor-Sánchez (2018, p. 74, tradução nossa) "Assim, por meio do *Schema.org* podem ser estruturados os conteúdos as páginas *Web* seguindo esquemas comuns baseados na categorização em campos e o controle de vocabulários que favorecem a recuperação da informação."

Apesar de ser criado para diversos domínios, em especial para o Web, o *Schema.org* tem sido usado na estruturação em *Linked Data* do *WorldCat*, catálogo que busca reunir diversas bibliotecas em um catálogo universal gerenciado pela OCLC.

Para realização do mapeamento do PROV para o *Schema.org* foram levadas em consideração as classes e das propriedades das duas ontologias. Entretanto, destaca-se que em alguns momentos não foram identificadas classes equivalentes em ambas ontologias, mas em alguns casos foram possíveis mapear equivalências entre classes e propriedades. Para o mapeamento, foram consideradas principalmente as classes *Thing*, CreativeWork e a tipologia *Book* do *Schema.org*.

Quadro 16 - PROV-O *crosswalk* Schema.org

PROV	Schema.org	Comentário
Classes		
prov:Entity	-	No Schema.org não foi localizada classe compatível
prov:Activity	Action	Os conceitos apresentados pelas classes são

		similares.
prov:Agent	Organization - Person	Os conceitos apresentados pelas classes são similares.
prov:Collection	Collection	Os conceitos apresentados pelas classes são similares.
prov:EmptyCollection	-	No Schema.org não foi localizada classe compatível
prov:Bundle	-	No Schema.org não foi localizada classe compatível
prov:Person	Person	Os conceitos apresentados pelas classes são similares.
prov:SoftwareAgent	-	No Schema.org não foi localizada classe compatível
prov:Organization	Organization	Os conceitos apresentados pelas classes são similares.
prov:Location	Place	Os conceitos apresentados pelas classes são similares.
prov:Influence	-	No Schema.org não foi localizada classe compatível
prov:EntityInfluence	-	No Schema.org não foi localizada classe compatível
prov:Usage	-	No Schema.org não foi localizada classe compatível
prov:Start	data - datatime	Os conceitos apresentados pelas classes são similares.
prov:End	data - datatime	Os conceitos apresentados pelas classes são similares.
prov:Derivation	workExample	Os conceitos apresentados pelas classes são similares.
prov:PrimarySource	isBasedOn	Os conceitos apresentados pelas classes são similares.
prov:Quotation	Quotation	Os conceitos apresentados pelas classes são similares.
prov:Revision	Review	Os conceitos apresentados pelas classes são similares.
prov:ActivityInfluence	-	No Schema.org não foi localizada classe compatível
prov:Generation	result	Os conceitos apresentados pelas classes são similares.
prov:Communication	CommunicateAction	Os conceitos apresentados pelas classes são similares.
prov:Invalidation	-	No Schema.org não foi localizada classe compatível
prov:AgentInfluence	-	No Schema.org não foi localizada classe compatível
prov:Attribution	-	No Schema.org não foi localizada classe compatível
prov:Association	-	No Schema.org não foi localizada classe compatível
prov:Plan	-	No Schema.org não foi localizada classe compatível
prov:Delegation	-	No Schema.org não foi localizada classe compatível
prov:InstantaneousEvent	-	No Schema.org não foi localizada classe compatível
prov:Role	-	No Schema.org não foi localizada classe compatível
prov:wasGeneratedBy	-	No Schema.org não foi localizada propriedade compatível
prov:wasDerivedFrom	-	No Schema.org não foi localizada propriedade compatível
prov:wasAttributedTo	-	No Schema.org não foi localizada propriedade

		compatível
prov:startedAtTime	dateCreated	Os conceitos apresentados pelas propriedades são similares.
	previousStartDate	
	startDate	
	startTime	
prov:used	-	No Schema.org não foi localizada propriedade compatível
prov:wasInformedBy	-	No Schema.org não foi localizada propriedade compatível
prov:endedAtTime	dissolutionDate	Os conceitos apresentados pelas propriedades são similares.
	endDate	
	dateDeleted	
	endTime	
prov:wasAssociatedWith	-	No Schema.org não foi localizada propriedade compatível
prov:actedOnBehalfOf	sourceOrganization	Os conceitos apresentados pelas propriedades são similares.
prov: alternateOf	-	No Schema.org não foi localizada propriedade compatível
prov: specializationOf	-	No Schema.org não foi localizada propriedade compatível
prov: generatedAtTime	-	No Schema.org não foi localizada propriedade compatível
prov: hadPrimarySource	-	No Schema.org não foi localizada propriedade compatível
prov: value	-	No Schema.org não foi localizada propriedade compatível
prov: wasQuotedFrom	citation	Os conceitos apresentados pelas propriedades são similares.
prov: wasRevisionOf	-	No Schema.org não foi localizada propriedade compatível
prov: invalidatedAtTime	-	No Schema.org não foi localizada propriedade compatível
prov: wasInvalidatedBy	-	No Schema.org não foi localizada propriedade compatível
prov: hadMember	-	No Schema.org não foi localizada propriedade compatível
prov: wasStartedBy	-	No Schema.org não foi localizada propriedade compatível
prov: wasEndedBy	-	No Schema.org não foi localizada propriedade compatível
prov:invalidated	-	No Schema.org não foi localizada propriedade compatível
prov:influenced	-	No Schema.org não foi localizada propriedade compatível
prov:atLocation	location	Os conceitos apresentados pelas propriedades são

	toLocation	similares.
prov:generated	-	No Schema.org não foi localizada propriedade compatível
prov:wasInfluencedBy	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedInfluence	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedGeneration	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedDerivation	workExample	Os conceitos apresentados pelas propriedades são similares.
	exampleOfWork	
	review	
prov:qualifiedPrimarySource	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedQuotation	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedRevision	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedAttribution	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedInvalidation	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedStart	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedUsage	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedCommunication	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedAssociation	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedEnd	-	No Schema.org não foi localizada propriedade compatível
prov:qualifiedDelegation	-	No Schema.org não foi localizada propriedade compatível
prov:influencer	-	No Schema.org não foi localizada propriedade compatível
prov:entity	-	No Schema.org não foi localizada propriedade compatível
prov:hadUsage	-	No Schema.org não foi localizada propriedade compatível
prov:hadGeneration	-	No Schema.org não foi localizada propriedade compatível
prov:activity	-	No Schema.org não foi localizada propriedade compatível
prov:agente	Agente	Os conceitos apresentados pelas propriedades são similares.
prov:hadPlan	-	No Schema.org não foi localizada propriedade

		compatível
prov:hadActivity	-	No Schema.org não foi localizada propriedade compatível
prov:atTime	startTime	Os conceitos apresentados pelas propriedades são similares.
	endTime	
prov:hadRole	-	No Schema.org não foi localizada propriedade compatível

Fonte: Elaborado pelo autor

No *crosswalk* entre o PROV-O para o *Schema.org*, foi possível mapear 14 classes e oito (8) propriedades. Assim, como nos outros padrões, em alguns casos houve mapeamentos de um-para-muitos e muitos-para-um.

5.5 PREMIS 3.0 Ontology

O *PREservation Metadata: Implementation Strategies* (PREMIS) é uma iniciativa de 2003 da *Online Computer Library Center* (OCLC) e pelo *Research Libraries Group* (RLG) e posteriormente passou a ser administrado pela *Library of Congress* (EUA). O PREMIS possui um conjunto básico de elementos de metadados de preservação e está baseado no modelo *Open Archival Information System* (OAIS). (ARAKAKI et al., 2017)

O PREMIS define cinco entidades: *Environment* (Suporte), *Object* (Objeto), *Event* (Evento), *Agent* (Agente), e *Rights Statement* (Declaração de direitos). Suportes podem ser descritos como Entidades Intelectuais e capturados e preservados no repositório como Representações, Arquivos e / ou *Bitstreams.* Suporte são ainda, Tecnologias (*software* ou *hardware*) de um Objeto Digital de alguma forma (por exemplo, renderização ou execução). Objeto (ou Objeto Digital) pode ser definida como uma unidade discreta de informação sujeita a preservação digital e usado como parte do processo de preservação. Evento é uma ação que envolve ou afeta pelo menos um objeto ou agente associado ou conhecido. Agente pode ser pessoa, organização ou programa / sistema de *software* associado a eventos na vida de um objeto ou com direitos associados a um objeto. Por fim, a Declaração de direitos é afirmação de um ou mais direitos ou permissões pertencentes a um objeto e / ou agente. (PREMIS…, 2015).

Nesse contexto, pensando no desenvolvimento das tecnologias, principalmente no que se refere à *Web* Semântica, a *Library of Congress* (EUA) desenvolveu o PREMIS 3.0 que

consiste na formalização semântica do dicionário de dados PREMIS 2.2 por meio de uma ontologia em OWL. (DI IORIO; CARON, 2016). Alguns estudos como Li e Sugimoto (2014, 2017, 2018) realizaram um mapeamento entre o PROV e o PREMIS, com o intuito de criar um modelo de proveniência no contexto da preservação digital.

No quadro 17, apresenta um *crosswalk* do PROV para o PREMIS.

Quadro 17 - *Crosswalk* PROV para PREMIS

PROV	PREMIS	Comentário
prov:Entity	Object	Os conceitos apresentados pelas classes são similares.
prov:Activity	Action	Os conceitos apresentados pelas classes são similares.
	Event	
prov:Agent	Agent	Os conceitos apresentados pelas classes são similares.
prov:Collection		Não foi localizada classe compatível
prov:EmptyCollection		Não foi localizada classe compatível
prov:Bundle	Representation	Os conceitos apresentados pelas classes são similares.
prov:Person	Person	Os conceitos apresentados pelas classes são similares.
prov:SoftwareAgent	SoftwareAgent	Os conceitos apresentados pelas classes são similares.
prov:Organization	Organization	Os conceitos apresentados pelas classes são similares.
prov:Location	StorageLocation	Os conceitos apresentados pelas classes são similares.
prov:Influence		Não foi localizada classe compatível
prov:EntityInfluence		Não foi localizada classe compatível
prov:Usage		Não foi localizada classe compatível
prov:Start		Não foi localizada classe compatível
prov:End		Não foi localizada classe compatível
prov:Derivation		Não foi localizada classe compatível
prov:PrimarySource		Não foi localizada classe compatível
prov:Quotation		Não foi localizada classe compatível
prov:Revision		Não foi localizada classe compatível
prov:ActivityInfluence		Não foi localizada classe compatível
prov:Generation		Não foi localizada classe compatível
prov:Communication		Não foi localizada classe compatível
prov:Invalidation		Não foi localizada classe compatível
prov:AgentInfluence		Não foi localizada classe compatível
prov:Attribution		Não foi localizada classe compatível
prov:Association		Não foi localizada classe compatível

prov:Plan		Não foi localizada classe compatível
prov:Delegation		Não foi localizada classe compatível
prov:InstantaneousEvent		Não foi localizada classe compatível
prov:Role		Não foi localizada classe compatível
prov:wasGeneratedBy	relationship	Os conceitos apresentados pelas classes são similares.
prov:wasDerivedFrom	version	Os conceitos apresentados pelas classes são similares.
prov:wasAttributedTo		Não foi localizada propriedade compatível
prov:startedAtTime		Não foi localizada propriedade compatível
prov:used		Não foi localizada propriedade compatível
prov:wasInformedBy		Não foi localizada propriedade compatível
prov:endedAtTime		Não foi localizada propriedade compatível
prov:wasAssociatedWith		Não foi localizada propriedade compatível
prov:actedOnBehalfOf		Não foi localizada propriedade compatível
prov: alternateOf		Não foi localizada propriedade compatível
prov: specializationOf		Não foi localizada propriedade compatível
prov: generatedAtTime		Não foi localizada propriedade compatível
prov: hadPrimarySource		Não foi localizada propriedade compatível
prov: value		Não foi localizada propriedade compatível
prov: wasQuotedFrom	citation	Os conceitos apresentados pelas classes são similares.
prov: wasRevisionOf		Não foi localizada propriedade compatível
prov: invalidatedAtTime		Não foi localizada propriedade compatível
prov: wasInvalidatedBy		Não foi localizada propriedade compatível
prov: hadMember		Não foi localizada propriedade compatível
prov: wasStartedBy	startDate	Os conceitos apresentados pelas classes são similares.
prov: wasEndedBy	endDate	Os conceitos apresentados pelas classes são similares.
prov:invalidated		Não foi localizada propriedade compatível
prov:influenced		Não foi localizada propriedade compatível
prov:atLocation		Não foi localizada propriedade compatível
prov:generated		Não foi localizada propriedade compatível
prov:wasInfluencedBy		Não foi localizada propriedade compatível
prov:qualifiedInfluence		Não foi localizada propriedade compatível
prov:qualifiedGeneration		Não foi localizada propriedade compatível
prov:qualifiedDerivation		Não foi localizada propriedade compatível
prov:qualifiedPrimarySource		Não foi localizada propriedade compatível
prov:qualifiedQuotation		Não foi localizada propriedade compatível
prov:qualifiedRevision		Não foi localizada propriedade compatível
prov:qualifiedAttribution		Não foi localizada propriedade compatível
prov:qualifiedInvalidation		Não foi localizada propriedade compatível
prov:qualifiedStart		Não foi localizada propriedade compatível
prov:qualifiedUsage		Não foi localizada propriedade compatível

prov:qualifiedCommunication		Não foi localizada propriedade compatível
prov:qualifiedAssociation		Não foi localizada propriedade compatível
prov:qualifiedEnd		Não foi localizada propriedade compatível
prov:qualifiedDelegation		Não foi localizada propriedade compatível
prov:influencer		Não foi localizada propriedade compatível
prov:entity		Não foi localizada propriedade compatível
prov:hadUsage		Não foi localizada propriedade compatível
prov:hadGeneration		Não foi localizada propriedade compatível
prov:activity	act	Os conceitos apresentados pelas classes são similares.
prov:agent		Não foi localizada propriedade compatível
prov:hadPlan		Não foi localizada propriedade compatível
prov:hadActivity		Não foi localizada propriedade compatível
prov:atTime		Não foi localizada propriedade compatível
prov:hadRole		Não foi localizada propriedade compatível

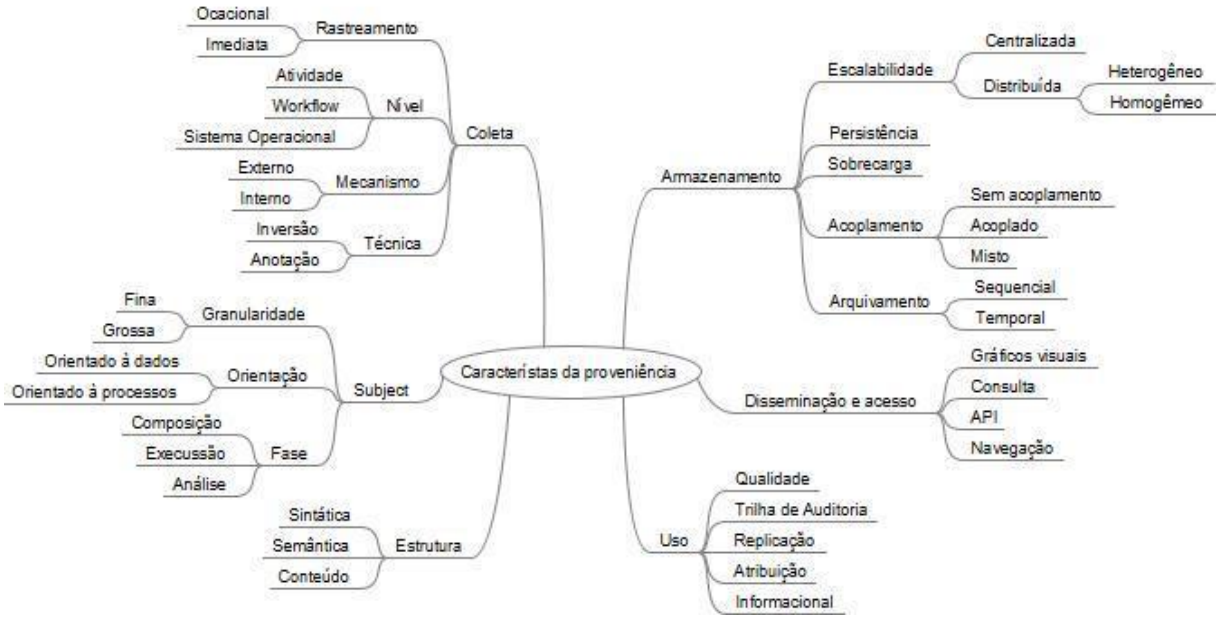
Fonte: Elaborado pelo autor

Na página de apresentação da ontologia do PREMIS, há informação que a ontologia teve base no PROV-O, entretanto, foram identificadas no mapeamento oito (8) classes e seis (6) propriedades utilizadas de fato.

6 DISCUSSÃO E ANÁLISE DOS DADOS

Neste capítulo tem como intuito de abordar a análise dos dados da pesquisa. Dessa forma, resgatando a Taxonomia das características da proveniência proposta nesta tese está baseada nos autores Simmhan, Plale e Gannon (2005) e Cruz, Campos e Mattoso (2009) e consiste em seis (6) categorias: Coleta, Representação, Subject, Armazenamento, Disseminação e acesso e Uso conforme apresentado na figura 10[5].

Figura 10 - Taxonomia das características da proveniência



Fonte: Elaborado pelo autor

Na coleta foram mantidas as subcategorias **Rastreamento**, **Nível** e **Mecanismo**. O Rastreamento pode ser dividido em **Ocasional** e **Imediato**. **Ocasional** calcula a proveniência de uma informação somente quando necessário, já em Imediato coleta a proveniência imediatamente à coleta dos dados. Já a subcategoria **Nível,** pode ser feita por ***workflow*** (fluxo de trabalho), **atividade** e **Sistema operacional**. ***Workflow*** é uma forma de coleta, obtenção e armazenamento de dados de proveniência. Nível de atividades requer a captura de cada serviço ou processo envolvido de coleta. A coleta por **sistema operacional**, a coleta

5 Para construção da figura foi utilizado o *Software* livre *FreeMind*

é realizada no nível do sistema operacional via API. No **mecanismo**, há duas formas **Interna** e **externa**.

A categoria **Representação** na taxonomia de Cruz, Campos e Mattoso (2009) pertencia à categoria captura como subcategoria denominada Technique. Entretanto, devido à complexidade da representação, optou-se em manter como categoria e com terminologia **Representação**. Como subcategoria da representação, optou-se em manter o Scheme (**método**), além de manter as subcategorias **Sintática**, **Semântica** e C**onteúdo**. O **Método** pode ser por **Inversão** ou **anotação**. A **inversão** é o processo de representação que pode ser **ah-hoc**, **manual** ou **automática**. A representação **Sintática** corresponde em como as informações vão ser descritas, como por exemplo, se serão adotadas linguagens de marcação. Já a representação **Semântica**, está atrelada ao uso de vocabulários controlados e ontologias e conteúdo que corresponde às informações contidas nos metadados.

Na categoria **Subject,** é estabelecida as formas de coleta e níveis de detalhamento, é definida pelas subcategorias **Orientação**, **Granularidade** e **Fase**. A **Orientação** pode ser por processo, ou seja, por etapas ou por dados. A **Granularidade** pode ser **fina** ou **grossa**, ou seja, quanto maior o nível de detalhamento e exaustividade pode ser considerada **Granularidade fina**, quanto menos detalhamento é considerada **Granularidade grossa**. **Fase** pode ser por três momento: **Composição prospectiva**, **Execução retrospectiva** e **Análise**. Na **Fase de composição**, incorpora as etapas que precisam ser seguidas para gerar um produto ou classe. Na F**ase de Execução**, coleta a procedência retrospectiva, os passos que foram executados, e as formas de execução utilizados para derivar algo. Na **Fase de Análise**, avalia os resultados e as mudanças realizadas.

Já na categoria Armazenamento, que estabelece questões de armazenamento das informações, possui como subcategorias: escalabilidade, sobrecarga, acoplamento, persistência e arquivamento. A escalabilidade corresponde à capacidade de gerenciar ou manipular informações. Sobrecarga corresponde ao limite de armazenamento. Acoplamento, dividido em três formas, sem acoplamento, acoplado e acoplamento misto. Sem acoplamento armazena informações em diversos repositórios. Acoplado, armazena a proveniência junto aos dados que a proveniência é registrada. Já o acoplamento solto, usa um esquema de armazenamento misto, que estão armazenados em um sistema de armazenamento, mas, separados logicamente. A persistência corresponde ao armazenamento não volátil da informação. O Arquivamento pode ser sequencial por

processos ou registro de alterações de tempo.

Por fim, na categoria Disseminação e Acesso pode ser por gráficos visuais, consultas e serviços API e navegação.

Na categoria **Uso**, prevalece a proposta de Simmhan, PLanle e Gannon (2005) com as subcategorias: **Qualidade de dados**, **Rrilha de Auditoria**, **Replicação, Atribuição** e **Informacional**. Na **Qualidade** pode ser utilizada para estimar a confiabilidade dos dados e transformação dos dados. Em **Trilha de auditoria**, utilizada para rastrear a trilha dos dados. Em **Replicação**, informações que permitem a repetição da derivação dos dados. Na **Atribuição**, a garantia dos direitos autorais e a propriedade dos dados. Por fim, em **Informacional**, a consulta nos metadados para descoberta de dados.

Com o levantamento e a análise dos metadados contidos nos padrões de metadados, foi possível identificar os elementos que apresentam características e podem ser usados para identificação da proveniência nos registros bibliográficos. Conforme destacado na metodologia, foi utilizado o método do *Crosswalk* de St Pierre e La plant (1999) para realizar as correspondências entre os padrões.

Para a análise desta pesquisa, foram extraídos todos os elementos do PROV-O e realizada o *Crosswalk*, conforme apresentado no quadro 18.

Quadro 18 - Visão geral do *Crosswalk*

PROV-O	MARC	DC	BIBFRAME	Schema.org	PREMIS
prov:Entity	-	dct:BibliographicResource dct:LicenseDocument dct:RightsStatement dct:PhysicalResource	-	-	Object
prov:Activity	583 - Action Note (R)	-	-	Action	Action Event
prov:Agent	100 - Main Entry - Personal Name (NR) 110 - Main Entry - Corporate Name (NR) 040 - Cataloging Source (NR) 700 - Added Entry - Personal Name (R) 710 - Added Entry - Corporate Name	dct:Agent	Agent	Organization Person	Agent
Classes	-	-			
prov:Collection	-	-	Collection	Collection	-
prov:EmptyCollection	-	-	-		
prov:Bundle	-	dct:ProvenanceStatement	-	-	Representation
prov:Person	100 - Main Entry - Personal Name (NR)	-	Person	Person	Person
prov:SoftwareAgent	-	-	-	-	SoftwareAgent

prov:Organization	110 - Main Entry - Corporate Name (NR) 040 - Cataloging Source (NR)	-	Organization	Organization	Organization
prov:Location	008 - Fixed-Length Data Elements-General Information (NR) - 15-17 - Place of publication, production, or execution 033 - Date/Time and Place of an Event (R) 260 - Publication, Distribution, etc. (Imprint) (R) 264 - Production, Publication, Distribution, Manufacture, and Copyright Notice (R) 370 - Associated Place (R) 518 - Date/Time and Place of an Event Note (R) 535 - Location of Originals/Duplicates Note (R) 544 - Location of Other Archival Materials Note (R) 852 - Location (R) 856 - Electronic Location and Access (R)	dct:Location dct:LocationPeriodOrJurisdiction	Local	Place	StorageLocation
Classe					
prov:Influence	-	-	-	-	-
prov:EntityInfluence	-	-	-	-	-
prov:Usage	-	-	UsageAndAccessPolicy AccessPolicy UsePolicy RetentionPolicy"	-	-
prov:Start	388 - Time Period of Creation (R) 033 - Date/Time and Place of an Event (R) 045 - Time Period of Content (NR) 046 - Special Coded Dates (R) 362 - Dates of Publication and/or Sequential Designation (R) 363 - Normalized Date and Sequential Designation (518 - Date/Time and Place of an Event Note (R)	-	creationDate	data - datatime	-
prov:End	005 - Date and Time of Latest Transaction Full |	-	-	data - datatime	-

	Concise				
prov:Derivation	251 - Version Information (R) 380 - Form of Work (R) 533 - Reproduction Note (R)	-	-	workExample exampleOfWork	-
prov:PrimarySource	534 - Original Version Note (R)	-	Source	isBasedOn	-
prov:Quotation	510 - Citation/References Note (R)	-	-	Quotation	-
prov:Revision	-	-	-	Review	-
prov:ActivityInfluence	-	-	-	-	-
prov:Generation	-	-	Generation GenerationProcess	result	-
prov:Communication	-	-	-	CommunicateAction	-
prov:Invalidation	-	-	-	-	-
prov:AgentInfluence	-	-	-	-	-
prov:Attribution	-	-	-	-	-
prov:Association	-	-	-	-	-
prov:Plan	-	dct:LinguisticSystem dct:MethodOfAccrual dct:MethodOfInstruction dct:Policy	-	-	-
prov:Delegation	-	-	-	-	-
prov:InstantaneousEvent	-	-	-	-	-
prov:Role	-	-	Role	-	-
	-	-	-	-	-
Propriedade	-	-	-	-	-
prov:wasGeneratedBy	883 - Machine-generated Metadata Provenance (R) - $a - Generation process (NR)	-	Generate	-	relationship
prov:wasDerivedFrom	380 - Form of Work (R) - $a - Form of work (R) 533 - Reproduction Note (R) - $a - Type of reproduction (NR)	dct:isFormatOf dct:references dct:source	derivativeOf ou derivedFrom	-	version
prov:wasAttributedTo	245 - Title Statement (NR) - $c - Statement of responsibility, etc. (NR)	dct:creator dct:contributor dct:publisher dct:rightsHolder	responsibilityStatement	-	-
prov:startedAtTime	388 - Time Period of Creation (R) - $a - Time period of creation term (R) 046 - Special Coded Dates (R) - $ k - Início ou data única criada (NR 362 - Dates of Publication and/or Sequential Designation (R) - $a - Dates of publication and/or sequential designation (NR) 518 - Date/Time and Place of an Event Note (R) - $ a - Data / hora e local de uma nota de	-	creationDate	dateCreated previousStartDate	-

	evento (NR) 518 - Date/Time and Place of an Event Note (R) - $ d - Data do evento (R) 883 - Machine-generated Metadata Provenance (R) - $d - Generation date (NR)				
prov:used	-	-	-	-	-
prov:wasInformedBy	-	-	-	-	-
prov:endedAtTime	046 - Special Coded Dates (R) - $ l - Data final criada (NR) 883 - Machine-generated Metadata Provenance (R)- $x - Validity end date (NR)	-	-	dissolutionDate endDate	-
prov:wasAssociatedWith	100 - Main Entry-Personal Name (NR) - $j - Attribution qualifier (R)	-	responsibilityStatement	-	-
prov:actedOnBehalfOf	533 - Reproduction Note (R) - $c - Agency responsible for reproduction (R)	-	-	sourceOrganization	-
	-	-	-	-	-
Propriedade	-	-	-	-	-
prov: alternateOf	-	dct:hasFormat dct:isFormatOf	-	-	-
prov: specializationOf	-	-	-	-	-
prov: generatedAtTime	-	dct:created dct:dateAccepted dct:dateCopyrighted dct:dateSubmitted dct:issued dct:modified	-	-	-
prov: hadPrimarySource	534 - Original Version Note (R) - $a - Type of reproduction (NR)	dct:source	Source	-	-
prov: value	-	-	-	-	-
prov: wasQuotedFrom	-	-	references – referencedBy	citation	citation
prov: wasRevisionOf	251 - Version Information (R) - $a - Version (R)	dct:isVersionOf	-	-	-
prov: invalidatedAtTime	-	-	-	-	-
prov: wasInvalidatedBy	-	-	-	-	-
prov: hadMember	583 - Action Note (R) - $k - Action agent (R)	-	-	-	-
prov: wasStartedBy	583 - Action Note (R) - $c - Time/date of action (R)	-	-	-	startDate
prov: wasEndedBy	-	-	-	-	endDate
prov:invalidated	-	-	-	-	-
prov:influenced	-	-	-	-	-
prov:atLocation	260 - Publication, Distribution, etc. (Imprint) (R) - $a - Place of	-	physicalLocation - place - sublocation	location toLocation	-

	publication, distribution, etc. (R)				
prov:generated	-	-	-	-	-
	-	-	-	-	-
Propriedade	-	-	-	-	-
prov:wasInfluencedBy	-	-	-	-	-
prov:qualifiedInfluence	-	-	-	-	-
prov:qualifiedGeneration	-	-	-	-	-
prov:qualifiedDerivation	251 - Version Information (R) - $a - Version (R) 533 - Reproduction Note (R - $a - Type of reproduction (NR)	-	-	workExample exampleOfWork	-
prov:qualifiedPrimarySource	-	-	-	-	-
prov:qualifiedQuotation	-	-	-	-	-
prov:qualifiedRevision	-	-	-	-	-
prov:qualifiedAttribution	-	-	-	-	-
prov:qualifiedInvalidation	-	-	-	-	-
prov:qualifiedStart	-	-	-	-	-
prov:qualifiedUsage	-	-	-	-	-
prov:qualifiedCommunication	-	-	-	-	-
prov:qualifiedAssociation	-	-	-	-	-
prov:qualifiedEnd	-	-	-	-	-
prov:qualifiedDelegation	-	-	-	-	-
prov:influencer	-	-	-	-	-
prov:entity	-	-	-	-	-
prov:hadUsage	-	-	-	-	-
prov:hadGeneration	-	-	-	-	-
prov:activity	583 - Action Note (R) - $a - Action (NR)	-	-	-	act
prov:agent	-	-	-	agent	-
prov:hadPlan	-	-	-	-	-
prov:hadActivity	-	-	-	-	-
prov:atTime	-	-	-	startTime endTime	-
prov:hadRole	-	-	-	-	-

Fonte: Elaborado pelo autor

Classes mapeadas:

- Duas (2) foram identificadas em todos os padrões mapeados: *Agent* e *Location*.
- Duas (2) classes do PROV foram mapeadas em 4 dos 5 padrões (*prov:Person; prov:Organization*).
- Três (3) classes foram mapeadas para 3 dos 5 padrões (*prov:Activity; prov:Start; prov:PrimarySource*).
- Sete (7) classes foram mapeadas entre 2 dos 5 padrões (*prov:Entity; prov:Collection; prov:Bundle; prov:End; prov:Derivation; prov:Quotation; prov:Generation*).
- Seis (6) classes foram mapeadas entre 1 dos 5 padrões (*prov:SoftwareAgent;*

prov:Usage; prov:Revision; prov:Communication; prov:Plan; prov:Role).

- Já dez (10) classes não foram mapeadas em nenhum dos 5 padrões (*prov:EmptyCollection; prov:Influence; prov:EntityInfluence; prov:ActivityInfluence; prov:Invalidation; prov:AgentInfluence; prov:Attribution; prov:Association; prov:Delegation; prov:InstantaneousEvent*).

Em relação às propriedades:

- Nenhuma propriedade foi mapeada nos 5 padrões analisados.
- Uma (1) propriedade do PROV foi mapeada em 4 dos 5 padrões (*prov:wasDerivedFrom*).
- Seis (6) propriedades foram mapeadas para 3 dos 5 padrões (*prov:wasGeneratedBy; prov:wasAttributedTo; prov:startedAtTime; prov: hadPrimarySource; prov: wasQuotedFrom; prov:atLocation*).
- Sete (7) propriedades foram mapeadas entre 2 dos 5 padrões (*prov:endedAtTime; prov:wasAssociatedWith; prov:actedOnBehalfOf; prov: wasRevisionOf; prov: wasStartedBy; prov:qualifiedDerivation; prov:activity*).
- Seis (6) propriedades foram mapeadas entre 1 dos 5 padrões (*prov: alternateOf; prov: generatedAtTime; prov: hadMember; prov: wasEndedBy; prov:agent; prov:atTime*).
- Trinta e uma (31) propriedades não foram mapeadas em nenhum dos 5 padrões (*prov:used; prov:wasInformedBy; Propriedade; prov: specializationOf; prov: value; prov: invalidatedAtTime; prov: wasInvalidatedBy; prov:invalidated; prov:influenced; prov:generated; prov:wasInfluencedBy; prov:qualifiedInfluence; prov:qualifiedGeneration; prov:qualifiedPrimarySource; prov:qualifiedQuotation; prov:qualifiedRevision; prov:qualifiedAttribution; prov:qualifiedInvalidation; prov:qualifiedStart; prov:qualifiedUsage; prov:qualifiedCommunication; prov:qualifiedAssociation; prov:qualifiedEnd; prov:qualifiedDelegation; prov:influencer; prov:entity; prov:hadUsage; prov:hadGeneration; prov:hadPlan; prov:hadActivity; prov:hadRole*).

7 CONSIDERAÇÕES FINAIS

Esta pesquisa teve como foco, discutir a importância da proveniência no domínio bibliográfico, principalmente no que se refere à publicação de dados abertos de ambientes digitais, promovendo assim o reuso de dados.

A questão norteadora da pesquisa foi "**qual a função dos metadados de proveniência nos registros bibliográficos em ambientes digitais?**" Dessa forma, para responder a essa questão foi proposto o objetivo geral **analisar a viabilidade do PROV, para garantia da proveniência dos dados nos registros bibliográficos, possibilitando a interoperabilidade de padrões de metadados e sistemas heterogêneos.** Entretanto, para alcançar a tal objetivo, foram traçados objetivos específicos para compor o objetivo geral e responder à questão de pesquisa proposta.

Dessa forma, no **capítulo dois,** foi apresentada uma revisão de literatura sobre a questão da Proveniência em diversos contextos, contemplando assim, o objetivo "**Realizar um levantamento do conceito de proveniência em diversos contextos**". Nesse sentido, foi possível verificar e discutir como a proveniência é abordada nos contextos da: Arquivologia, Museologia, Artes, Jornalismo, Ciências no âmbito geral, Ambiente digital, Preservação digital, Computação, Reuso de dados, Dados científicos e Web Semântica.

Nos contextos da Arquivologia, Museologia e Artes, a proveniência tem dupla função, além da confiabilidade das informações e relatar a origem do documento e quem criou, pode ser utilizada para organização dos fundos, das coleções e dos acervos.

A Computação tem apresentado diversas contribuições na garantia da coleta, armazenamento disseminação e uso da proveniência, principalmente no que diz respeito ao reuso de dados, ambientes digitais e *Web* Semântica. Entre elas destaca-se a Taxonomia das características da proveniência em *workflows*, que é ressaltado a importância da proveniência em diversos âmbitos.

Em cada contexto apresentado, mostrou-se a importância da proveniência na garantia da confiabilidade e da integridade dos dados. Dessa forma, foi possível identificar que a proveniência está relacionada principalmente em três aspectos (Entidade, Atividade e Agente) conforme descrito por Lebo, Sahoo e McGuinnes (2013).

A Entidade, ou seja, informações de proveniência do objeto, coisa, abstrata ou física. Exemplos de informações de proveniência de uma entidade é a fonte primária, a coleção,

acervo ou fundo que pertence, entre outras. As Atividades podem ser consideradas também as ações que uma entidade sofreu ao longo de sua vida. Podem ser consideradas informações de proveniência de uma atividade, derivações, atualizações, publicações, entre outras. Por fim, o Agente corresponde a uma pessoa, organização ou *software* que realiza alguma ação ou possui uma entidade. Exemplo de informação de proveniência é quem criou, ou realizou alguma alteração, ou publicou alguma entidade, entre outras.

Assim como na questão da publicação dos dados, a proveniência é elemento fundamental para o reuso de dados, confiança e também na compreensão. Concomitantemente, no contexto da *Web* Semântica, a proveniência apresentou contribuições principalmente para o desenvolvimento camada da confiança, conforme apresentado por Moreau e Groth (2013).

Por conseguinte, no capítulo três, foram apresentados alguns documentos fundamentais da família PROV para dar subsídios para validação da hipótese, contemplando assim o objetivo "**Apresentar as características e as relações das especificações da família PROV**". Nesse contexto, foi abordado uma breve contextualização das iniciativas para representação de proveniência, em especial o PROV. Além disso, foi apresentado neste capítulo documentos da família PROV: PROV-PRIMER, PROV-DM e PROV-O.

O PROV-PRIMER é um documento introdutório que aborda diversos conceitos básicos do PROV que são utilizados em outros documentos, isto é, pode ser considerado como um ponto de partida para quem quer iniciar e entender o PROV.

Posteriormente, foi abordado o PROV-DM que é um modelo de dados para representação da proveniência, ou seja, uma abstração dos elementos e dos relacionamentos de proveniência que podem ser descritos em um sistema de informação. Foi modelado em UML e é a base para todos os outros documentos da família PROV.

Já o PROV-O define as classes e as propriedades utilizando a linguagem OWL2 com base no PROV-DM, além de estabelecer as relações entre as entidades e as propriedades, restrições e inferências da ontologia. Ele apresenta três categorias de uso. O primeiro nível com as três classes principais (Entidade, Atividade e Agente). O segundo nível denominado expandido, apresenta subclasses para classes principais entre outras subclasses. Já o terceiro nível, qualificado, apresenta diversas outras classes e propriedades, além de deixar em aberto um possível tipo de qualificação de algumas propriedades.

Após o referencial teórico levantado nos capítulos dois, três e quatro foi possível

entender a questão da proveniência, em que contexto os metadados de proveniência estão relacionados, e ainda, as possibilidades de aplicação do PROV no domínio bibliográfico.

Sendo assim, o levantamento do contexto de uso da proveniência fez questionar que tipologia de metadados está inserido as informações de proveniência. Dessa forma, o **capítulo quatro**, apresentou uma discussão sobre os metadados e suas tipologias, contemplando o objetivo "**Estabelecer as características dos metadados de proveniência e verificar a relação com as demais tipologias de metadados**". Nessa análise é possível observar que não há um consenso na literatura referente às tipologias dos metadados.

A abordagem para os metadados descritivos apresenta-se como a mais consensual na literatura, com relação aos seus objetivos, categorias e elementos que os representam. Os metadados administrativos, entretanto, não há consenso na literatura quanto suas subcategorias. Entre as tipologias encontradas como subcategoria dos metadados administrativos foram: Direitos, Acesso, Estruturais, Meta-metadata, Preservação, Proveniência, Técnico e Uso. Apenas as tipologias Meta-metadata e Proveniência não foram consideradas como uma tipologia independente. Todas as demais tipologias, Direitos, Acesso, Estruturais, Preservação, Técnico e Uso foram consideradas por pelo menos um autor, como tipologia independente. Já os metadados de autenticidade e *Markup Languages*, foram citados por um autor, como tipologias independentes.

A tipologia dos metadados administrativos foi citadas apenas por um autor como subtipologia dos metadados administrativos (POMERANTZ, 2015). Em outros casos, alguns autores comentaram que há informações de proveniência, mas não colocaram como tipologia ou subtipologia, apenas que são metadados que pertencem a tipologia dos metadados administrativos.

Dessa forma, a o objetivo "**Identificar a proveniência nos padrões de metadados aplicados ao domínio bibliográfico (MARC21, Dublin Core, BIBFRAME, PREMIS e Schema.org).**" foi apresentado no capítulo cinco. Nesse sentido, foi contextualizado os principais ambientes digitais no domínio bibliográfico como catálogos, OPACs, Repositórios, Sistema de descoberta pensando em possíveis aplicações da questão da proveniência. Assim foram elencados os principais padrões de metadados para descrição desses itens, como o MARC21, *Dublin Core*, BIBFRAME, PREMIS e *Schema.org*.

Dessa forma, cada padrão foi mapeamento individualmente, com base no método *Crosswalk*. Conforme o método, foi realizado a equivalência entre classes, propriedades,

campos e subcampos, posteriormente a identificação das terminologias e as definições dos termos. Posteriormente, foi realizado o mapeamento semântico das classes e das propriedades que resultou em mapeamentos um-para-um, um-para-muitos, muitos-para-um e um-para-nenhum.

Após a realização do mapeamento individual, o capítulo seis apresentou a análise dos dados. Dessa forma, ao decorrer da pesquisa, observou grande relação entre as classes das categorias apresentadas por Simmhan, Plale, Gannon (2005) e Cruz, Campos e Mattoso (2009) a partir de uma releitura das taxinomias propostas, com o tripé da proveniência (entidade, atividade e agente). Ou seja, tanto a entidade, a atividade e o agente, possui as características da taxonomia proposta, pois, nas três, haverá um momento para coleta e essa coleta poderá ser categorizada por diversas formas como, rastreamento, nível, tipos de mecanismo e técnicas de coleta. Subjetc, com subcategorias como granularidade, orientação (objeto ou processo) e fase. A questão da estrutura também é fundamental, assim foram categorizados a estrutura sintática, semântica e de conteúdo. Em algum dado momento, a proveniência precisa ser armazenada, possibilitando como subcategorias a escalabilidade, persistência, sobrecarga, acoplamento e arquivamento. Outro aspecto é a disseminação e acesso que foram subcategorizados em visualização por gráficos, consultas, API e navegação. Por fim, no uso das informações pode ser analisado a questão da qualidade, trilha de auditoria, replicação, atribuição e informacional.

Já em relação ao mapeamento, foi criado um quadro geral com todos os mapeamentos. Considera-se que as classes básicas do PROV são poucas para representação da complexidade das informações de proveniência, sendo necessário o uso do PROV estendido e qualificado para melhor representar a complexidade da representação da proveniência no domínio bibliográfico. Destaca-se que esse cenário, com registros mais exaustivos, amplia as possibilidades de acesso aos recursos informacionais seria o ideal para representação, entretanto, poucas bibliotecas teriam condições de gerar registros tão completos, ou por falta de recursos humanos ou tecnológicos. Uma alternativa para solucionar esse problema, seria pensar em registrar alguns metadados de forma automática.

Com o levantamento bibliográfico, junto à taxonomia das características da proveniência e mapeamento dos padrões foi possível validar a hipótese principal desta pesquisa que consistia em “**que o modelo PROV-O pode ser aplicado ao domínio bibliográfico para a representação da proveniência em registros bibliográficos, permitindo**

a descrição da origem, das ações e dos envolvidos na construção e alteração de registros em ambientes digitais", pois, foi ressaltada a importância dos metadados de proveniência, caracterizado e defendido como uma subtipologia dos metadados administrativos, além da validação da possibilidade de uso do PROV no domínio bibliográfico. Destaca-se que o PROV apresenta diversos metadados que podem ser incorporados nos registros bibliográficos para complementar a descrição dos recursos informacionais.

Nesse contexto, considera-se que a tese **"os registros bibliográficos necessitam de metadados referentes à proveniência dos dados para a preservação da integridade e da persistência, como forma na garantia da confiabilidade das informações em ambientes digitais"** foi confirmada.

Como trabalhos futuros destaca-se:

- Realizar o aprofundamento teórico de como a proveniência pode contribuir para as camadas da *Web* Semântica, principalmente no que diz respeito à camada de confiança;
- Estudar a questão da proveniência em padrões específicos para descrição de pessoas e entidades coletivas;
- Verificar se a taxonomia criada, poderia resultar em um fluxo de trabalho para proveniência no domínio bibliográfico;
- Discutir e ampliar elementos fundamentais de proveniência para o domínio bibliográfico;
- Discutir a questão da proveniência na certificação de repositórios digitais confiáveis.

O estudo sobre a proveniência, a patir do levantamento bibliográfico, mostrou que a temática no Brasil é incipiente e carece de pesquisas teóricas e iniciativas de cunho prático/profissional. Tal conclusão, revela a necessidade de um cuidado especial no planejamento do sistema, na definição dos metadados que irão compor o registro bibliográfico do recurso informacional. Isso irá refletir na definição de quais informações são necessárias aos usuários, para o acesso e a visualização e, por último, e não menos importante, na definição de quais informações são necessárias para gestão e curadoria do sistema.

REFERÊNCIAS

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. 132 f. f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília/SP, 2010. Disponível em: <http://repositorio.unesp.br/handle/11449/103361>.

ALVES, Rachel Cristina Vesu; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.

ALVES, Rachel Cristina Vesu; SIMIONATO, Ana Carolina; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Aspectos de granularidade na representação da informação no universo bibliográfico. 2012, Rio de Janeiro. *Anais*... Rio de Janeiro: [s.n.], 2012.

ANDRADE, Morgana; BAPTISTA, Ana Alice. The Use of Application Profiles and Metadata Schemas by Digital Repositories: Findings from a Survey. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS 2015, set. 2015, [S.l.]: Dublin Core Metadata Initiative (DCMI), set. 2015. p. 146–157. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/37855>. Acesso em: 31 dez. 2018.

ARAKAKI, Felipe Augusto; GALEFFI, Luiz Felipe; et al. BIBFRAME: tendência para a representação bibliográfica na web. **RBBD. Revista Brasileira de Biblioteconomia e Documentação**, v. 13, n. 0, p. 2231–2249, 23 dez. 2017. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/995>. Acesso em: 31 dez. 2018.

ARAKAKI, Felipe Augusto. **Linked Data: ligação de dados bibliográficos**. 2016. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP), 2016. Disponível em: <https://repositorio.unesp.br/handle/11449/147979>.

ARAKAKI, Felipe Augusto; GONÇALEZ, Paula Regina Ventura Amorim; et al. Web Semântica e preservação digital: o padrão de metadados PREMIS na proposta do Linked Data. In: COLÓQUIO DE DADOS, METADADOS E WEB SEMÂNTICA, 4 dez. 2017, São Carlos. *Anais*... São Carlos: UFSCar, 4 dez. 2017. Disponível em: <https://cdmws.isci.com.br/ocs/index.php/cdmws/home/paper/view/35>. Acesso em: 31 dez. 2018.

ARAKAKI, Felipe Augusto; ALVES, Rachel Cristina Vesu; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Dublin Core: State of Art (1995 to 2015). **Informação & Sociedade: Estudos**, v. 28, n. 2, 28 ago. 2018. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/38012>. Acesso em: 31 dez. 2018.

ARAÚJO, Paula Carina de et al. Serviço de descoberta: implantação no Sistema de Bibliotecas (SiBi) da Universidade Federal do Paraná (UFPR). In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 2015, Florianópolis. *Anais*... Florianópolis: Anais do XXVI Congresso Brasileiro de Biblioteconomia e Documentação (CBBD/2015), 2015.

ARQUIVO NACIONAL (BRASIL). **Dicionário brasileiro de terminologia arquivística**. . Rio de Janeiro: Arquivo Nacional, 2015. Disponível em: <http://www.arquivonacional.gov.br/images/pdf/Dicion_Term_Arquiv.pdf>.

ASSUMPÇÃO, Fabrício Silva. **Modelo Para a Publicação de Dados de Autoridade Como Linked Data**. 2018. 208 f. (Tese de doutorado). Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP). Disponível em: <http:// hdl.handle.net/11449/152759>. Acesso em: 16 jan. 2019.

ASSUMPÇÃO, Fabrício Silva; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Representação no domínio bibliográfico: um olhar sobre os Formatos MARC 21. **Perspectivas em Ciência da Informação**, v. 20, n. 1, p. 54–74, 30 mar. 2015. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2054>. Acesso em: 31 dez. 2018.

BACA, Murtha (Org.). **Introduction to metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/>.

BAKER, Thomas. A Grammar of Dublin Core. **D-Lib Magazine**, v. 6, n. 10, 2000. Disponível em: <http://www.dlib.org/dlib/october00/baker/10baker.html>. Acesso em: 4 dez. 2016.

_____. Libraries, Languages of Description, and Linked Data: A Dublin Core Perspective. **Library Hi Tech**, v. 30, n. 1, p. 116–133, 2 mar. 2012. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/07378831211213256>. Acesso em: 4 dez. 2016.

_____. **Library linked data incubator group final report**. Disponível em: <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>. Acesso em: 4 dez. 2016.

BALBY, Claudia Negrão. **Estudos de uso de catálogos on-line (OPACs):** revisão metodológica e aplicação da técnica de análise de log de transações a um OPAC de biblioteca universitária brasileira. 2002. Tese (Doutorado em Ciência da Informação) – Universidade de São Paulo, São Paulo, 2002. Disponível em: <http://bdpi.usp.br/item/001278951>.

_____. Formatos de intercâmbio de registros bibliográficos: conceitos básicos. **Cadernos da FFC**, v. 4, n. 1, p. 29–34, 1995.

BAPTISTA, Ana Alice. **Desafios à comunidade ibero-americana de metadados em repositórios digitais para maximização da interoperabilidade**. [S.l.]: Editora UTFPR, 2017. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/49022>. Acesso em: 31 dez. 2018.

BARBOSA, Alice Príncipe. **Novos rumos da catalogação**. Rio de Janeiro: BNG/Brasilart, 1978.

BASTOS, Flávia Maria. **A interação do usuário com catálogos bibliográficos on-line: investigação a partir da teoria fundamentada**. 2013. 255 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília/SP, 2013. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/bastos_fm_do_mar.pdf>.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. **Scientific american**, v. 284, n. 5, p. 28–37, 2001.

BISSET ALVAREZ, Edgar. **Sistemas de recomendação para bibliotecas universitárias: um aporte teórico da arquitetura da informação**. 2017. 182 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília/SP, 2017. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alvarez_eb_do.pdf>.

BIVAR, Bárbara et al. Uma Comparação entre os Modelos de Proveniência OPM e PROV. **Proocedings of BRESCi**, 2013.

BORKO, Harold. Information science: what is it? **American documentation**, v. 19, n. 1, p. 3–5, 1968.

BREEDING, Marshall. Library Services Platforms: A Maturing Genre of Products. **Library Technology Reports**, v. 51, n. 4, 1 maio 2015. Disponível em: <https://librarytechnology.org/repository/item.pl?id=21299>. Acesso em: 31 dez. 2018.

_____. **Robots in Academic Libraries**. [S.l.]: IGI Global, 2013. Disponível em: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-3938-6>. Acesso em: 31 dez. 2018. (Advances in Library and Information Science).

BUNEMAN, Peter; KHANNA, Sanjeev; TAN, Wang-Chiew. Data provenance: Some basic issues. 2000, [S.l.]: Springer, 2000. p. 87–93.

CAMARGO, Liriane Soares de Araújo de; VIDOTTI, Silvana Aparecida Borsetti Gregorio. **Arquitetura da informação**. Rio de Janeiro: Grupo Gen - LTC, 2011. . Acesso em: 31 dez. 2018.

CARVALHO, José et al. Auditoria ISO 16363 a repositórios institucionais. **Cadernos BAD**, v. 0, n. 2, p. 29–39, 2014. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/1175>. Acesso em: 31 dez. 2018.

CASTRO, Fabiano Ferreira de; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Elementos de interoperabilidade na perspectiva da catalogação descritiva. **Informação & Sociedade: Estudos**, v. 24, n. 3, 2014. Disponível em: <http://periodicos.ufpb.br/index.php/ies/article/view/16660>. Acesso em: 31 dez. 2018.

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization–a study of methodology part I. **D-Lib magazine**, v. 12, n. 6, p. 3, 2006.

CHOWDHURY, G. G.; CHOWDHURY, Sudatta. **Organizing information**. London: Facet, 2007.

CICCARESE, Paolo et al. PAV Ontology: Provenance, Authoring and Versioning. **Journal of Biomedical Semantics**, v. 4, n. 1, p. 37, 2013. Disponível em: <http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-37>. Acesso em: 30 dez. 2018.

CONSELHO NACIONAL DE ARQUIVOS. **Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos**. . [S.l.]: Arquivo Nacional Rio de Janeiro. Disponível em: <http://www.conarq.gov.br/images/publicacoes_textos/diretrizes_rdc_arq.pdf>. , 2015

CRUZ, Sérgio Manuel Serra da; CAMPOS, Maria Luiza M.; MATTOSO, Marta. Towards a taxonomy of provenance in scientific workflow management systems. 2009, [S.l.]: IEEE, 2009. p. 259–266.

DI IORIO, Angela; CARON, Bertrand. PREMIS 3.0 Ontology: Improving Semantic Interoperability of Preservation Metadata. In: INTERNATIONAL CONFERENCE ON DIGITAL PRESERVATION, 2016, [S.l: s.n.], 2016. p. 32–36. Disponível em: <https://www.loc.gov/standards/premis/pif/2016/iPresDiIoria_Caron_iPRES_2016_paper_150.pdf>.

DURANTI, Luciana et al. **InterPARES Trust Terminologia**. Disponível em: <http://arstweb.clayton.edu/interlex/pt/>. Acesso em: 30 dez. 2018.

ECKERT, Kai. **Metadata Provenance in Europeana and the Semantic Web**. 2012. Humboldt-Universität zu Berlin, Philosophische Fakultät I, 2012. Disponível em: <https://edoc.hu-berlin.de/bitstream/handle/18452/2727/332.pdf?sequence=1&isAllowed=y>.

ECKERT, Kai; GARIJO, Daniel; PANZER, Michael. Extending DCAM for Metadata Provenance. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2011, The Hague. *Anais*... The Hague: [s.n.], 2011. p. 12–25. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3621>. Acesso em: 31 dez. 2018.

FERREIRA, Margarida M. **MARC 21**. Marília: UNESP, 2002.

FERREIRA, Norma Sandra de Almeida. As Pesquisas Denominadas “Estado Da Arte”. **Educação & Sociedade**, v. 23, n. 79, ago. 2002. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-73302002000300013&lng=pt&nrm=iso&tlng=pt>. Acesso em: 4 dez. 2016.

FONSECA, Fernanda Maria Lobo da; ANDRADE, Leila Cristina Rodrigues de. SERVIÇO DE DESCOBERTA: CONSIDERAÇÕES SOBRE A IMPLANTAÇÃO NA REDE DE BIBLIOTECAS DA UNIVERSIDADE DO ESTADO DO RIO DE JANEIRO–UERJ. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 2014, Belo Horizonte, MG. *Anais*... Belo Horizonte, MG: [s.n.], 2014. p. 1–17.

GARIJO, Daniel; ECKERT, Kai. **Dublin Core to PROV Mapping**. Disponível em: <https://www.w3.org/TR/prov-dc/>. Acesso em: 30 dez. 2018.

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

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. [S.l: s.n.], 2010. Disponível em: <http://alltitles.ebrary.com/Doc?id=10824884>. Acesso em: 4 dez. 2016.

GIL, Yolanda; MILES, Simon. **PROV Model Primer**. Disponível em: <https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/>. Acesso em: 30 dez. 2018.

GIL, Yolanda et al. **Provenance XG Final Report**. Disponível em: <https://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/#What_is_provenance>. Acesso em: 30 dez. 2018.

GILLILAND, Anne J. Setting the Stage. In: BACA, Murtha (Org.). . **Introd. Metadata**. 1. ed. Los Angeles: Getty Research Institute, 1999. . Disponível em: <http://www.getty.edu/publications/intrometadata/>.

_____. Setting the Stage. In: BACA, Murtha (Org.). . **Introd. Metadata**. 2. ed. Los Angeles: Getty Research Institute, 2008. . Disponível em: <http://www.getty.edu/publications/intrometadata/>.

_____. Setting the Stage. In: BACA, Murtha (Org.). . **Introd. Metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. . Disponível em: <http://www.getty.edu/publications/intrometadata/>.

GLUSHKO, Robert J. **The Discipline of Organizing**. 1 ed. ed. Massachusetts, EUA: The MIT Press, 2013. Disponível em: <http://site.ebrary.com/id/10841924>. Acesso em: 4 dez. 2016. (MIT Press).

GONZALES, Brighid M. Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. **Information Technology and Libraries**, v. 33, n. 4, 18 dez. 2014. Disponível em: <http://ejournals.bc.edu/ojs/index.php/ital/article/view/5631>. Acesso em: 31 dez. 2018.

GREENBERG, Jane. A quantitative categorical analysis of metadata elements in image-applicable metadata schemas. **Journal of the American Society for Information Science and Technology**, v. 52, n. 11, p. 917–924, 2001.

HANSEN, Jytte; ANDRESEN, Leif. **AC - Administrative Components**. . [S.l.]: DCMI. Disponível em: <dublincore.org/groups/admin/AdminComp_final_June_2003.doc>. , 2003

HAYNES, David. **Metadata for information management and retrieval**. [S.l.]: Facet Publishing, 2004. v. 1.

_____. **Metadata for Information Management and Retrieval: Understanding metadata and its use**. [S.l.]: Facet Publishing, 2018.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **IFLA -- Guidelines for Online Public Access Catalogue (OPAC) Displays (2005)**. Disponível em: <https://www.ifla.org/FR/publications/ifla-series-on-bibliographic-control-27>. Acesso em: 31 dez. 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 16363**. . [S.l: s.n.]. Disponível em:

<http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/65/56510.html>. Acesso em: 31 dez. 2018. , 2012

JEFFERY, K. et al. A 3-Layer model for metadata. 2013, [S.l: s.n.], 2013. p. 3–5. Disponível em: <http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/199/199>.

JOUDREY, Daniel N.; TAYLOR, Arlene G.; WISSER, Katherine M. **The organization of information**. Fourth edition ed. Santa Barbara, California: Libraries Unlimited, 2018. (Library and information science text series).

KOVARI, Jason; FOLSOM, Steven; YOUNES, Rebecca. Towards a BIBFRAME Implementation: The Bibliotek-o Framework. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2 dez. 2017, Whashington. **Anais**... Whashington: DCMI, 2 dez. 2017. p. 52–61. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3854>. Acesso em: 31 dez. 2018.

KUMAR, Sharma; UJJAL, Marjit; UTPAL, Biswas. Exposing MARC 21 Format for Bibliographic Data Ass Linked Data with Provenance. **Journal of Library Metadata**, v. 13, n. 2–3, p. 212–229, jul. 2013. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/19386389.2013.826076>. Acesso em: 30 dez. 2018.

LEBO, Timothy; SAHOO, Satya; MCGUINNESS, Deborah. **PROV-O: The PROV Ontology**. Disponível em: <https://www.w3.org/TR/2013/REC-prov-o-20130430/>. Acesso em: 30 dez. 2018.

LI, Chunqiu; SUGIMOTO, Shigeo. Provenance Description of Metadata Application Profiles for Long-Term Maintenance of Metadata Schemas. **Journal of Documentation**, v. 74, n. 1, p. 36–61, 8 jan. 2018. Disponível em: <http://www.emeraldinsight.com/doi/10.1108/JD-03-2017-0042>. Acesso em: 30 dez. 2018.

_____. Provenance description of metadata using PROV with PREMIS for long-term use of metadata. 2014, [S.l: s.n.], 2014. p. 147–156.

_____. Provenance description of metadata vocabularies for the long-term maintenance of metadata. **Journal of Data and Information Science**, v. 2, n. 2, p. 41–55, 2017.

LIBRARY OF CONGRESS. **Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services**. . [S.l.]: Library of Congress. Disponível em: <https://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf>. , 2012

_____. **MARC 21 Format for Bibliographic Data: 883: Machine-generated Metadata Provenance**. Disponível em: <https://www.loc.gov/marc/bibliographic/bd883.html>. Acesso em: 31 dez. 2018.

_____. **MARC 21 Format for Bibliographic Data : Introduction (Network Development and MARC Standards Office, Library of Congress)**. Disponível em: <https://www.loc.gov/marc/bibliographic/bdintro.html>. Acesso em: 31 dez. 2018.

_____. **Overview of the BIBFRAME 2.0 Model**. webpage. Disponível em: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>. Acesso em: 31 dez. 2018.

LIBRARY OF CONGRESS; PREMIS EDITORIAL COMMITTEE. **PREMIS Data Dictionary for Preservation Metadata, Version 3.0 (Library of Congress)**. webpage. Disponível em: <http://www.loc.gov/standards/premis/v3/>. Acesso em: 30 dez. 2018.

LIU, Jia. **Metadata and its applications in the digital library: approaches and practices**. [S.l.]: Libraries Unlimited Westport, CT, 2007.

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. **Data on the Web Best Practices**. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 30 dez. 2018.

MACEDO, S. Ascensão de. 'Proveniência' na terminografia arquivística de língua portuguesa: prospeção e visualização de (dis)similaridades em termos e definições. **Revista Ibero-Americana de Ciência da Informação**, v. 11, n. 2, p. 388–409, 28 maio 2018. Disponível em: <http://periodicos.unb.br/ojs311/index.php/RICI/article/view/8334>. Acesso em: 30 dez. 2018.

MARANHÃO, Ana Maria Neves. A SELEÇÃO DE UM SERVIÇO DE DESCOBERTA NA WEB A EXPERIÊNCIA DA PUC-RIO. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 2012, Gramado. **Anais**... Gramado: [s.n.], 2012. p. 1–16. Disponível em: <https://s3.amazonaws.com/academia.edu.documents/46005663/snbu_2012.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1546264535&Signature=vFs4%2BjWWWPYTwS0iGEFGaIJQCUs%3D&response-content-disposition=inline%3B%20filename%3DA_SELECAO_DE_UM_SERVICO_DE_DESCOBERTA_NA.pdf>.

MÉNDEZ RODRÍGUEZ, Eva Ma. **Metadatos y recuperación de información**. Gijón, Asturias: Ediciones Trea, 2002. (Biblioteconomía y administración cultural, 66).

MEY, Eliane Serrão Alves. **Introdução à Catalogação**. Brasília: Briquet de Lemos, 1995.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofoletti. **Catalogação no plural**. [S.l.]: Briquet de Lemos, 2009.

MILLER, Steven J. **Metadata for digital collections**. New York: Neal-Schuman Publishers, 2011. (How-to-do-it manuals, no. 179).

MOREAU, Luc et al. The Open Provenance Model Core Specification (v1.1). **Future Generation Computer Systems**, v. 27, n. 6, p. 743–756, jun. 2011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X10001275>. Acesso em: 30 dez. 2018.

MOREAU, Luc; GROTH, Paul. Provenance: an introduction to prov. **Synthesis Lectures on the Semantic Web: Theory and Technology**, v. 3, n. 4, p. 1–129, 2013. Disponível em: <https://www.morganclaypool.com/doi/abs/10.2200/S00528ED1V01Y201308WBE007>.

MOREAU, Luc; MISSIER, Paolo. **PROV-DM: The PROV Data Model**. Disponível em: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>. Acesso em: 30 dez. 2018a.

_____. **PROV-N: The Provenance Notation**. Disponível em: <https://www.w3.org/TR/2013/REC-prov-n-20130430/>. Acesso em: 30 dez. 2018b.

NELSON, David; TURNEY, Linda. What's in a word? : Rethinking facet headings in a discovery service. **Information Technology and Libraries**, v. 34, n. 2, p. 76–91, 14 jun. 2015. Disponível em: <http://ejournals.bc.edu/ojs/index.php/ital/article/view/5629>. Acesso em: 31 dez. 2018.

NIES, Tom De. **Constraints of the PROV Data Model**. Disponível em: <https://www.w3.org/TR/prov-constraints/>. Acesso em: 30 dez. 2018.

NILSSON, Mikael; BAKER, Thomas; JOHNSTON, Pete. **DCMI: The Singapore Framework for Dublin Core Application Profiles**. Disponível em: <http://dublincore.org/documents/singapore-framework/>. Acesso em: 31 dez. 2018.

NÓBREGA-THERRIEN, Sílvia Maria; THERRIEN, Jacques. Trabalhos científicos e o estado da questão. **Estudos em Avaliação Educacional**, v. 15, n. 30, p. 5, 30 dez. 2004. Disponível em: <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2148>. Acesso em: 4 dez. 2016.

PAVÃO, Caterina Marta Groposo. **Comportamento de busca e recuperação da informação em serviços de descoberta em rede no contexto acadêmico**. 2014. 219 f. Tese (Doutorado em Comunicação) – Universidade Federal do Rio Grande do Sul, Rio Grande do SUl, 2014. Disponível em: <https://lume.ufrgs.br/handle/10183/96705>. Acesso em: 31 dez. 2018.

PEARCE-MOSES, Richard; BATY, Laurie A. **A glossary of archival and records terminology**. [S.l.]: Society of American Archivists Chicago, IL, 2005. v. 2013. Disponível em: <http://www.chismechick.com/wp-content/uploads/2017/08/SAA-Glossary-2005.pdf>.

PIERRE, Margaret St; LAPLANT, William P. Issues in crosswalking content metadata standards. **Information standards quarterly**, v. 11, n. 1, p. 01–16, 1999. Acesso em: 28 out. 2016.

POMERANTZ, Jeffrey. **Metadata**. Cambridge, Massachusetts ; London, England: The MIT Press, 2015. (The MIT Press essential knowledge series).

POWELL, Andy et al. **DCMI: DCMI Abstract Model**. Disponível em: <http://dublincore.org/documents/2007/06/04/abstract-model/>. Acesso em: 31 dez. 2018.

PREMIS Data Dictionary for Preservation Metadata, Version 3.0. . [S.l.]: Library of Congress. Disponível em: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>. , 2015

PREMIS EDITORIAL COMMITTEE. PREMIS Data Dictionary for Preservation Metadata, Version 3.0. p. 283, 2015. Disponível em: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>.

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - UNESP. **Linhas de Pesquisa PPGCI - Unesp Campus de Marília**. Disponível em: <http://www.marilia.unesp.br/#!/pos-graduacao/mestrado-e-doutorado/ciencia-da-informacao/linhas-de-pesquisa/>. Acesso em: 30 dez. 2018.

RESEARCH LIBRARIES GROUP - OCLC. **Trusted Digital Repositories: Attributes and Responsibilities**. . [S.l.]: RLG-OCLC Report. Disponível em: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>. Acesso em: 8 ago. 2018. , 2002

REZENDE, Laura Vilela Rodrigues; CRUZ-RIASCOS, Sonia Sonia Aguiar; HOTT, Daniela Francescutti Martins. Em busca de repositórios digitais confiáveis no Brasil: análise da infraestrutura organizacional conforme a norma ISO 16363/2012. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 11, n. 0, 30 nov. 2017. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1390>. Acesso em: 31 dez. 2018.

RILEY, Jenn. **Understanding Metadata**. NISO Press: National Information Standards Organization (U.S.), 2004.

_____. **Understanding Metadata: what is metadata, and what is it for?** . [S.l.]: National Information Standards Organization (NISO). Disponível em: <http://www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf>. , 2017

ROA-MARTÍNEZ, Sandra Milena; VIDOTTI, Silvana Aparecida Borsetti Gregorio; PASTOR-SÁNCHEZ, Juan Antonio. Marcado semántico enriquecido para programas de posgrado en Latinoamérica. **Perspectivas em Ciência da Informação**, v. 23, n. 3, p. 67–88, set. 2018. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362018000300067&lng=es&tlng=es>. Acesso em: 31 dez. 2018.

SANCHEZ, Fernanda Alves [UNESP. **Encontrabilidade da informação em repositórios digitais**. 2018. 238 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília (SP), 2018. Disponível em: <https://repositorio.unesp.br/handle/11449/154348>. Acesso em: 31 dez. 2018.

SANKARI, R Lakshmi et al. A STUDY ON THE USE OF ONLINE PUBLIC ACCESS CATALOGUE (OPAC) BY STUDENTS AND FACULTY MEMBERS OF UNNAMALAI INSTITUTE OF TECHNOLOGY IN KOVILPATTI (TAMIL NADU). **International Journal of Library and Information Studies**, v. 3, p. 10, 2013. Disponível em: <http://www.ijlis.org/img/2013_Vol_3_Issue_1/17-26.pdf>.

SANTAREM SEGUNDO, José Eduardo [UNESP. **Representação iterativa**. 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marilia, 2010. Disponível em: <https://repositorio.unesp.br/handle/11449/103346>. Acesso em: 31 dez. 2018.

SANTOS, Henrique Machado dos; FLORES, Daniel. Repositórios digitais confiáveis para documentos arquivísticos: ponderações sobre a preservação em longo prazo. **Perspectivas em Ciência da Informação**, v. 20, n. 2, p. 198–218, 30 jun. 2015. Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2341>. Acesso em: 31 dez. 2018.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Redes informacionais como ambientes colaborativos e de empoderamento: a catalogação em foco. In: GUIMARÃES, José Augusto Chaves; FUJITA, Mariângela Spotti Lopes (Orgs.). **Ensino e pesquisa em biblioteconomia no Brasil**: a emergência de um novo olhar. Marília: Cultura acadêmica, 2008, p. 155-171

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; PEREIRA, Ana Maria. **Catalogação: breve história e contemporaneidade**. Niterói (RJ): Intertexto, 2014.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; SANT'ANA, Ricardo César Gonçalves. Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, v. 42, n. 2, p. 199–209, 2013.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41–62, 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, v. 21, n. 2, p. 90, 20 dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>. Acesso em: 15 ago. 2017.

SCHEMA.ORG. **about page - schema.org**. Disponível em: <https://schema.org/docs/about.html>. Acesso em: 31 dez. 2018.

SENSO, José A.; ROSA PIÑERO, Antonio De la. 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/>.

SILVA, Eduardo Graziosi; BOCCATO, Vera Regina Casari. Avaliação do uso de catálogos coletivos de bibliotecas universitárias pela perspectiva sociocognitiva do usuário. **Transinformação**, v. 24, n. 1, 2012.

SILVA, Marcel Ferrante; LIMA, Gercina Ângela Borém de Oliveira. Avaliação de usabilidade em interface de busca com navegação facetada e busca por palavra-chave. **Pesquisa Brasileira em Ciência da Informação**, v. 8, n. 1, p. 1–16, 2015. Disponível em: <https://www.researchgate.net/profile/Gercina_Lima/publication/324994656_Avaliacao_de_usabilidade_em_interface_de_busca_com_navegacao_facetada_e_busca_por_palavra-chave/links/5af0c455aca272bf42542219/Avaliacao-de-usabilidade-em-interface-de-busca-com-navegacao-facetada-e-busca-por-palavra-chave.pdf>.

SIMIONATO, Ana Carolina. **Modelagem conceitual DILAM: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital**. 2015. 200 f. f. Tese – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília/SP, 2015. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/simionato_ac_do_mar.pdf>.

SIMMHAN, Yogesh L.; PLALE, Beth; GANNON, Dennis. A survey of data provenance techniques. **Computer Science Department, Indiana University, Bloomington IN**, v. 47405, p. 69, 2005.

TAYLOR, Arlene G.; JOUDREY, Daniel N. **The organization of information**. 3rd ed ed. Westport, Conn: Libraries Unlimited, 2009. (Library and information science text series).

THANUSKODI, S. Use of Online Public Access Catalogue at Annamalai University Library. **International Journal of Information Science**, v. 2, n. 6, p. 70–74, 1 dez. 2012. Disponível em: <http://article.sapub.org/10.5923.j.ijis.20120206.01.html>. Acesso em: 31 dez. 2018.

TORINO, Emanuelle. **Repositórios digitais**. [S.l.]: EDUTFPR, 2017. Disponível em: <http://repositorio.utfpr.edu.br:8080/jspui/handle/1/2495>. Acesso em: 31 dez. 2018. (Repositórios digitais: teoria e prática).

VAUGHAN, Jason. Chapter 1: Web Scale Discovery What and Why? **Library Technology Reports**, v. 47, n. 1, p. 5–11, 5 jan. 2011. Disponível em: <https://journals.ala.org/index.php/ltr/article/view/4380>. Acesso em: 31 dez. 2018.

_____. Investigations into Library Web-Scale Discovery Services. **Information Technology and Libraries**, v. 31, n. 1, p. 32, 1 mar. 2012. Disponível em: <http://ejournals.bc.edu/ojs/index.php/ital/article/view/1916>. Acesso em: 31 dez. 2018.

VELLUCCI, Sherry L. Metadata. **Annual review of information science and technology (ARIST)**, v. 33, p. 187–222, 1998.

VIDOTTI, Silvana Ap Borsetti Gregorio et al. Coleta automática para povoamento de repositórios digitais: conversão de registros utilizando XSLT. **Tendências da Pesquisa Brasileira em Ciência da Informação**, 2017. Disponível em: <http://200.20.0.78/repositorios/handle/123456789/3590>.

WOODLEY, Mary S.; CLEMENT, Gail; WINN, Pete. **DCMI Glossary: using Dublin Core**. Disponível em: <http://www.dublincore.org/documents/usageguide/glossary/>. Acesso em: 30 dez. 2018.

ZENG, Marcia Lei; QIN, Jian. **Metadata**. New York: Neal-Schuman Publishers, 2008.

_____. **Metadata**. Second edition, UK edition ed. London: fp, facet publishing, 2016.