



FACULDADE DE **CIÊNCIA DA INFORMAÇÃO**

UNIVERSIDADE DE BRASÍLIA

Faculdade de Ciência da Informação

Curso de Graduação em Biblioteconomia

RDA Oficial e proveniência de dados:
desafios e relações baseadas no modelo PROV-O

Brasília

2024

Vítor Fernandes de Faria

RDA Oficial e proveniência de dados:
desafios e relações baseadas no modelo PROV-O

Monografia apresentada como parte das exigências para obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Ciência da Informação da Universidade de Brasília
Orientador: Prof. Dr. Felipe Augusto Arakaki

Brasília
2024

V845r Faria, Vítor Fernandes de
RDA Oficial e proveniência de dados : desafios e relações baseadas
no modelo PROV-O / Vítor Fernandes de Faria ; orientador Felipe Augusto
Arakaki. -- Brasília, 2024.

63 f. : il. color.

Monografia (Graduação - Biblioteconomia) -- Universidade
de Brasília, 2024.

Orientador: Prof. Dr. Felipe Augusto Arakaki

1. Proveniência de dados. 2. *Resource Description and Access*
(RDA). 3. Ontologia PROV (PROV-O). 4. Catalogação. I. Arakaki, Felipe
Augusto, orient. II. Título.

FOLHA DE APROVAÇÃO

Título: RDA Oficial e proveniência de dados: desafios e relações baseadas no modelo PROV-O.

Autor(a): Vítor Fernandes de Faria

Monografia apresentada em **20 de setembro de 2024** à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Orientador(a) (FCI/UnB): Dr. Felipe Augusto Arakaki

Membro Interno (FCI/UnB): Dra. Fernanda Passini Moreno

Membro Externo (UFSCar/IBICT): Dra. Ana Carolina Simionato Arakaki



Documento assinado eletronicamente por **Fernanda Passini Moreno, Membro do Colegiado do Curso de Biblioteconomia da Faculdade de Ciência da Informação**, em 24/10/2024, às 14:12, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Felipe Augusto Arakaki, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 24/10/2024, às 14:26, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Ana Carolina Simionato Arakaki, Usuário Externo**, em 28/10/2024, às 10:39, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **11936827** e o código CRC **70E0D734**.

AGRADECIMENTOS

A Deus por Sua infinita bondade e misericórdia, por me dar uma segunda chance e por sempre cuidar de mim e guiar-me ao longo dessa jornada. Sem Ele, nada disso seria possível.

À minha namorada e futura esposa, Heloísa, a quem agradeço do fundo do meu coração por estar ao meu lado em todos os momentos. Seu apoio, paciência e amor me ajudaram a chegar até aqui e a concluir essa etapa.

Aos meus pais, por todo o amor, apoio e confiança que sempre depositaram em mim. Obrigado por tudo que vocês me proporcionaram e por sempre acreditarem em mim.

À minha irmã, meu sincero agradecimento por sua amizade, carinho e por sempre estar ao meu lado, sempre me apoiando.

Aos meus amigos da igreja, que sempre me apoiaram e me ajudaram a chegar até aqui.

A minha chefe e amiga, Adelaide por sempre me apoiar e ajudar.

Ao meu amigo, João Lucas, com quem enfrentei todos os desafios e dificuldades ao longo das disciplinas na UnB. Agradeço por sua parceria nessa jornada.

Ao professor Felipe Arakaki, por ter aceitado ser meu orientador e por toda a paciência, dedicação e orientação ao longo desse trabalho. Sua ajuda foi fundamental para que eu pudesse desenvolver e concluir este estudo.

RESUMO

Esse estudo tem como objetivo investigar a integração entre o padrão *Resource Description and Access* (RDA) e as práticas de proveniência de dados, seguindo as diretrizes da ontologia PROV do *World Wide Web Consortium* (W3C), com foco na otimização da interação entre esses dois sistemas em ambientes digitais, onde a precisão e a confiabilidade das informações são essenciais. A pesquisa é justificada pela necessidade de aprimorar as práticas de catalogação e gestão de dados, garantindo maior transparência, rastreabilidade e autenticidade das informações. O objetivo principal é identificar e avaliar os elementos da RDA que contribuem para a gestão da proveniência de dados, analisando como esses elementos se alinham com as recomendações do, de forma a assegurar a integridade das informações bibliográficas e culturais. A pesquisa abrange a exploração da estrutura e dos princípios da RDA, investigando a definição e gestão de metadados relacionados à proveniência, bem como a comparação entre a RDA e a ontologia PROV para identificar semelhanças, diferenças e complementaridades. A metodologia adotada é qualitativa e exploratória, utilizando o método *Crosswalk* para mapear e comparar os elementos de metadados dos padrões, visando promover a interoperabilidade entre eles. A coleta de dados foi realizada por meio de uma revisão bibliográfica, que incluiu normas e recomendações do e do RDA Toolkit, além de artigos acadêmicos. A análise focou na correspondência entre os elementos dos padrões, com atenção aos desafios enfrentados, como o acesso limitado ao RDA Toolkit, que restringiu a uma análise mais ampla dos elementos. Os principais resultados indicam que 63,15% dos elementos analisados apresentaram correspondência direta com os elementos do PROV-O, sugerindo uma boa compatibilidade entre os padrões RDA e PROV-O, enquanto 26,31% dos elementos não apresentaram correspondência direta, e 10,54% não apresentaram nenhuma correspondência, deixando lacunas nos aspectos de proveniência pela RDA. Conclui-se que a RDA possui uma aderência parcial à ontologia PROV, com alguns aspectos de proveniência não totalmente cobertos, indicando a necessidade de uma análise mais aprofundada para verificar a compatibilidade total entre os padrões, além de sugerir que futuras pesquisas explorem as *Qualified Classes and Properties* do PROV-O, que não foram abordadas nesse estudo.

Palavras-chave: Proveniência de dados. *Resource Description and Access* (RDA). Ontologia PROV (PROV-O). Catalogação.

ABSTRACT

This study aims to investigate the integration between the Resource Description and Access (RDA) standard and data provenance practices, following the guidelines of the PROV ontology of the World Wide Web Consortium (W3C), focusing on optimizing the interaction between these two systems in digital environments, where the accuracy and reliability of information are essential. The research is justified by the need to improve cataloging and data management practices, ensuring greater transparency, traceability, and authenticity of information. The main objective is to identify and evaluate the RDA elements that contribute to the management of data provenance, analyzing how these elements align with W3C recommendations to ensure the integrity of bibliographic and cultural information. The research covers the exploration of the structure and principles of RDA, investigating the definition and management of provenance-related metadata, as well as the comparison between RDA and the PROV ontology to identify similarities, differences, and complementarities. The methodology adopted is qualitative and exploratory, using the Crosswalk method to map and compare metadata elements of the standards, aiming to promote interoperability between them. Data collection was carried out through a literature review, which included W3C standards and recommendations and the RDA Toolkit, in addition to academic articles. The analysis focused on the correspondence between the elements of the standards, paying attention to the challenges faced, such as limited access to the RDA Toolkit, which restricted a broader analysis of the elements. The main results indicate that 63.15% of the analyzed elements showed direct correspondence with the elements of PROV-O, suggesting good compatibility between the RDA and PROV-O standards, while 26.31% of the elements did not present direct correspondence, and 10.54% showed no correspondence, leaving gaps in the provenance aspects covered by RDA. It is concluded that RDA has partial adherence to the PROV ontology, with some provenance aspects not fully covered, indicating the need for further analysis to verify full compatibility between the standards, in addition to suggesting that future research explores the Qualified Classes and Properties of PROV-O, which were not addressed in this study.

Keywords: Data provenance. Resource Description and Access (RDA). PROV ontology (PROV-O). Cataloging.

LISTA DE ILUSTRAÇÕES

Gráfico 1 - Grau de correspondência entre RDA e PROV-O.....	56
Gráfico 2 - Tipos de <i>Crosswalk</i> entre RDA e PROV-O.....	57

LISTA DE TABELAS

Tabela 1 - Elementos selecionados do RDA Toolkit.	27
Tabela 2 - <i>Starting Point Terms</i> do PROV-O.	33
Tabela 3 - <i>Expanded Terms</i> do PROV-O.....	35
Tabela 4 - <i>Qualified Classes and Properties</i> do PROV-O.....	38
Tabela 5 - Comparação da RDA com as classes do PROV-O <i>Starting Point Terms</i>	48
Tabela 6 - Comparação da RDA com as propriedades do PROV-O <i>Starting Point Terms</i>	49
Tabela 7 - Comparação da RDA com as classes do PROV-O <i>Expanded Terms</i>	51
Tabela 8 - Comparação da RDA com as propriedades do PROV-O <i>Expanded Terms</i>	53

LISTA DE ABREVIATURAS E SIGLAS

AACR2	<i>Anglo-American Cataloguing Rules 2</i>
FRAD	<i>Functional Requirements for Authority Data</i>
FRBR	<i>Functional Requirements for Bibliographic Records</i>
IFLA	<i>International Federation of Library Associations and Institutions</i>
JSC	<i>Joint Steering Committee for Development of RDA</i>
OWL	<i>Web Ontology Language</i>
PROV	<i>Provenance</i>
PROV-O	<i>Provenance Ontology</i>
RDA	<i>Resource Description and Access</i>
W3C	<i>World Wide Web Consortium</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Problema	14
1.2 Justificativa	15
1.3 Objetivos	15
1.3.1 Objetivo geral	15
1.3.2 Objetivos específicos	16
1.4 Estrutura do Trabalho.....	16
2 REFERENCIAL TEÓRICO	19
2.1 RDA	19
2.1.1 Histórico da RDA	19
2.1.2 Estrutura da RDA	22
2.2 PROV	29
2.2.1 Conceitos de proveniência	29
2.2.2 Família PROV.....	31
2.2.3 Estrutura da Ontologia PROV (PROV-O).....	32
3 METODOLOGIA.....	43
3.1 Estrutura Metodológica	43
3.1.1 Abordagem da pesquisa	43
3.1.2 Método <i>Crosswalk</i>	43
3.1.3 Coleta de dados	44
3.1.4 Procedimentos de análise	45
3.1.5 Instrumentos norteadores	45
4 ANÁLISE DE DADOS.....	47
5 CONSIDERAÇÕES FINAIS.....	58
REFERÊNCIAS.....	61

1 INTRODUÇÃO

A crescente complexidade na organização e rastreamento de informações bibliográficas em ambientes digitais trouxe à tona a necessidade de padrões mais robustos e flexíveis, capazes de lidar com a proveniência de dados e a interoperabilidade de metadados. Nesse cenário, a *Resource Description and Access* (RDA) surge como um marco essencial, não apenas como um padrão de catalogação, mas como um sistema que facilita a rastreabilidade e a confiabilidade dos dados.

A integração da RDA com modelos como o PROV-O, uma ontologia desenvolvida pelo W3C para a gestão de proveniência, indica uma convergência entre a organização de metadados e o rastreamento da linhagem dos dados em múltiplos contextos tecnológicos (OLIVER, 2021).

A evolução no gerenciamento de dados bibliográficos e de patrimônio cultural foi significativamente impactada pela implementação do *Resource Description and Access* (RDA), especialmente após as reformulações trazidas pelo Projeto 3R. Publicado originalmente em 2010, a RDA passou por uma ampla revisão iniciada em 2017, no âmbito do Projeto 3R, que modernizou e alinhou o padrão às necessidades contemporâneas de catalogação e gestão de metadados (OLIVER, 2021).

Essa revisão incluiu mudanças substanciais no RDA Toolkit, com atualizações que abrangeram desde a modernização da interface do usuário até a melhoria das funcionalidades de edição e tradução. O objetivo central dessas alterações foi adaptar a RDA para uma maior internacionalização e sua aplicação em ambientes digitais. Além disso, o conteúdo da RDA foi atualizado para se alinhar ao mais recente modelo conceitual bibliográfico da *International Federation of Library Associations and Institutions* (IFLA), assegurando sua conformidade com padrões globais de descrição de recursos e acesso (OLIVER, 2021).

A RDA, que teve sua origem das Regras de Catalogação Anglo-Americanas (AACR2), desenvolveu-se em um padrão reconhecido internacionalmente. Foi criado para ser aplicável em uma ampla variedade de contextos bibliográficos, podendo ser ajustado para atender às exigências particulares de diversas comunidades, incluindo aquelas que não estão inseridas nos contextos tradicionais de bibliotecas. Sua abordagem mais inclusiva e global marca uma evolução em relação à visão majoritariamente anglo-americana de sua versão anterior,

permitindo sua implementação em diversos contextos linguísticos, culturais e tecnológicos (OLIVER, 2021).

A respeito da incorporação da RDA nas metodologias de rastreamento de dados, enfrentam-se desafios e surgem oportunidades. Os desafios consistem em ajustar a RDA, que é meticulosamente estruturado e embasado teoricamente nos modelos conceituais bibliográficos da IFLA, para atender às exigências variáveis e muitas vezes intrincadas da proveniência dos dados. Esse esforço implica em lidar com as sutilezas do registro de dados, assegurando a correta formação dos metadados e sua gestão eficaz em diversas plataformas tecnológicas, um objetivo que a RDA visa facilitar (OLIVER, 2021).

As oportunidades são muitas ao promover a flexibilidade da RDA e sua orientação para tarefas centradas no usuário como 'encontrar', 'identificar', 'selecionar', 'obter' e 'explorar'. A abordagem focada no usuário é fundamental no contexto da proveniência de dados, onde entender a origem, o contexto e a história dos dados servem para garantir sua confiabilidade e integridade. Além disso, o design da RDA para o ambiente digital e sua compatibilidade com uma variedade de esquemas de codificação o coloca como uma ferramenta versátil no gerenciamento de dados de proveniência em paisagens digitais diversas e interconectadas (OLIVER, 2021).

Em essência, segundo Oliver (2021), a integração da RDA nas práticas de proveniência de dados não é apenas um alinhamento técnico, mas uma harmonização conceitual, reunindo a abordagem estruturada e centrada no usuário da RDA com os requisitos complexos de rastreamento e manutenção da linhagem de dados. Essa síntese promete melhorar a qualidade dos dados, a interoperabilidade e o engajamento do usuário em várias aplicações.

No panorama da era digital, a gestão da proveniência de dados emergiu como um desafio fundamental. Confrontados com a crescente complexidade e volume de dados, os especialistas da área da Tecnologia da Informação e Biblioteconomia buscaram soluções para assegurar a autenticidade e a confiabilidade das informações. Nesse contexto, a ontologia PROV, desenvolvido pelo *World Wide Web Consortium (W3C)*, oferece um paradigma robusto para a representação e gestão da proveniência de dados em ambientes digitais (ARAKAKI, 2020).

A origem da ontologia PROV pode ser definida até a formação de um grupo de trabalho específico no W3C, dedicado exclusivamente às questões de proveniência de dados, o qual era composto por especialistas de diversos campos. Esse grupo empenhou-se na elaboração de um conjunto de padrões e recomendações que moldariam o futuro da gestão da proveniência. O fruto deste esforço colaborativo foi a criação da família PROV, um conjunto de documentos e recomendações que detalham um modelo abrangente para a representação da proveniência de dados (ARAKAKI, 2020).

Este novo modelo além de incorporar os princípios de outros padrões de proveniência, também os expandiu, estabelecendo uma base sólida para a avaliação da proveniência de dados, independente do modelo específico utilizado. Com sua abordagem ampla e flexível, o PROV se posicionou como um elemento chave na gestão de metadados de proveniência, adaptável a uma ampla gama de contextos e aplicações digitais (ARAKAKI, 2020).

Assim, segundo Arakaki (2020), a ontologia PROV não é apenas um conjunto de recomendações técnicas, o qual representa a concretização de um esforço coletivo para enfrentar os desafios impostos pela era digital na gestão de dados. Sua adoção e implementação contínua em diferentes domínios evidenciam seu papel relevante na garantia da transparência, rastreabilidade e confiabilidade das informações digitais, servindo como uma demonstração da contínua evolução na área de gestão de dados e metadados.

1.1 Problema

A RDA tem sido de grande importância na catalogação e na gestão de recursos bibliográficos. Contudo, sua integração com as práticas de proveniência de dados, cruciais para assegurar a integridade e autenticidade das informações, apresenta desafios. A proveniência de dados, no que se refere ao rastreamento da origem, histórico e autenticidade dos dados, é essencial para garantir sua confiabilidade e usabilidade em diversas aplicações.

No entanto, devido à complexidade da RDA, especialmente após as alterações significativas do Projeto 3R, são levantadas questões sobre como esse padrão pode ser efetivamente adaptado para suportar a proveniência de dados, preservando ao mesmo tempo a integridade e a riqueza das informações bibliográficas.

A lacuna existente entre os princípios de catalogação da RDA e as necessidades específicas das práticas de proveniência de dados cria um terreno fértil para a pesquisa, onde

explorar esses assuntos podem proporcionar percepções valiosas que aprimoram a gestão de informações. Nesse contexto, o principal questionamento que este trabalho busca responder é: qual a aderência da RDA Oficial à ontologia PROV?

1.2 Justificativa

Esse trabalho é justificado pela necessidade emergente de entender e otimizar a interação entre a RDA e as práticas de proveniência de dados. Em um ambiente onde a precisão e a confiabilidade dos dados são cada vez mais importantes, especialmente em contextos digitais, a integração eficaz da RDA com a proveniência de dados se torna essencial.

Ao abordar os desafios dessa integração, este trabalho busca contribuir para o aprimoramento das práticas de catalogação e gestão de dados, assegurando que a riqueza de informações fornecidas pela RDA seja adequada para melhorar a transparência, a rastreabilidade e a autenticidade dos dados.

A justificativa social do estudo sobre a integração do padrão RDA com as práticas de proveniência de dados em ambientes digitais é fundamental. Essa integração tem o potencial de aumentar a transparência e a confiabilidade das informações bibliográficas para os usuários finais, como estudantes, pesquisadores e o público geral, ao melhorar a acessibilidade e a usabilidade dos dados, o estudo contribui para uma sociedade mais informada, facilitando o acesso a informações precisas e confiáveis.

A motivação para o trabalho se deu através da crescente preocupação com a precisão e rastreabilidade das informações em ambientes digitais. O interesse na integração de padrões internacionais de catalogação, como a RDA com a ontologia PROV, reflete a necessidade de garantir a autenticidade e integridade dos dados, especialmente no contexto atual, em que a transparência dos dados é cada vez mais essencial.

1.3 Objetivos

1.3.1 Objetivo geral

A RDA Oficial é o foco deste estudo, que tem como objetivo identificar e avaliar os elementos que contribuem para a gestão da proveniência de dados. Este estudo explora como a RDA incorpora e se alinha com as recomendações do *World Wide Web Consortium* (W3C)

sobre proveniência de dados no ambiente digital, especialmente em relação à documentação do PROV, um padrão norteador da área.

O objetivo é investigar os metadados específicos dentro da RDA que indicam a proveniência de dados, avaliando como esses elementos facilitam a rastreabilidade, a transparência e a integridade das informações bibliográficas e patrimoniais culturais em contextos digitais. O trabalho busca compreender a extensão e a eficácia com que a RDA adota os princípios de proveniência de dados, contribuindo assim para o desenvolvimento de práticas de catalogação e gestão de informações mais confiáveis.

1.3.2 Objetivos específicos

- a. Explorar e compreender a RDA a partir do estudo da estrutura, princípios e elementos da RDA, com ênfase nos aspectos relacionados à proveniência de dados, investigando como a RDA define e gerencia metadados que contribuem para a rastreabilidade e a integridade dos dados.
- b. Identificar a família PROV de proveniência de dados, conforme recomendado pelo W3C, para entender suas diretrizes é abordado a documentação e gestão da proveniência em ambientes digitais.
- c. Comparar a RDA e a ontologia PROV, focando na identificação das semelhanças, diferenças e potenciais complementaridades entre as duas estruturas, visando entender como a RDA se alinha ou diverge das recomendações da ontologia PROV em termos de gestão da proveniência de dados.

1.4 Estrutura do Trabalho

Seção 1: Introdução

Apresenta o problema central do estudo, discutindo as dificuldades de integrar a RDA com as práticas de proveniência de dados e a necessidade de alinhar esses padrões para assegurar a integridade e autenticidade dos dados. Justifica a importância da pesquisa e define os objetivos geral e específicos do estudo, que visam avaliar como a RDA incorpora as recomendações do W3C sobre proveniência de dados.

Seção 2: Referencial Teórico

Apresenta a evolução da RDA, desde suas origens nas Regras Anglo-Americanas de Catalogação (AACR) até sua revisão e modernização através do Projeto 3R. Detalha os conceitos fundamentais e as tecnologias necessárias para a implementação da RDA, bem como os conceitos básicos de proveniência de dados conforme definidos pela ontologia PROV do W3C. Descreve os documentos e recomendações que compõem a família PROV, fornecendo uma base para a gestão de proveniência em ambientes digitais, e estabelece a estrutura detalhada da ontologia PROV.

Seção 3: Metodologia

Descreve os métodos e procedimentos adotados para uma análise rigorosa e detalhada da integração entre o padrão RDA e as práticas de proveniência de dados recomendadas pela ontologia PROV do W3C. A pesquisa adotou uma abordagem qualitativa, dada a complexidade e a natureza exploratória do tema, permitindo uma análise aprofundada das características, relações e implicações dessa integração. O método *Crosswalk* foi empregado para realizar a comparação detalhada entre os elementos dos padrões RDA e PROV, visando mapear as correspondências, identificar as semelhanças e diferenças, e avaliar a interoperabilidade entre esses esquemas de metadados. Também integra a análise dos diferentes graus de equivalência entre os elementos mapeados, a fim de verificar a aderência da RDA aos princípios de proveniência de dados estabelecidos pelo PROV-O. Além disso, foram considerados os desafios e limitações encontrados no uso do RDA Toolkit, que impactaram o processo de mapeamento.

Seção 4: Análise e resultados

Analisa de forma comparativa os elementos da RDA e as classes e propriedades da Ontologia PROV para avaliar a correspondência dos elementos RDA aos do PROV-O. Busca se há ou não equivalência entre os elementos e se há correspondentes, seja diretamente ou parcialmente, ainda destaca os desafios na integração dos conceitos de proveniência de dados mais propostos pelo PROV-O na estrutura da RDA. Engloba também os resultados e comparações a fim de verificar a aderência da RDA Oficial à ontologia PROV.

Seção 5: Considerações

Apresenta as conclusões do estudo após mapear e comparar os elementos dos dois padrões através do método *Crosswalk*, assim como resgata os objetivos apontando quais foram alcançados. Mostra os resultados e o alinhamento que há entre os dois padrões, com uma boa parte dos elementos da RDA correspondendo diretamente aos elementos do PROV-O. Adicionalmente, foram discutidos os desafios enfrentados durante a pesquisa, e se conclui com uma proposta para futuros trabalhos, com a sugestão de uma análise detalhada das *Qualified Classes and Properties do PROV-O* e de possíveis melhorias para a acessibilidade e usabilidade do RDA Toolkit.

2 REFERENCIAL TEÓRICO

A evolução das normas e padrões de catalogação reflete uma busca constante por precisão, consistência e interoperabilidade em ambientes cada vez mais digitais. Nesse cenário, conceitos como a descrição de recursos e a proveniência de dados ganham importância fundamental, especialmente quando se trata de garantir a rastreabilidade e a integridade da informação.

As normas contemporâneas, como aquelas adotadas no campo da Biblioteconomia, trazem à tona a necessidade de ferramentas que não só organizem e descrevam, mas que também assegurem a confiabilidade dos registros ao longo do tempo. Nesse contexto, surgem novos desafios e oportunidades para a gestão da informação, culminando na convergência entre estruturas convencionais e sistemas inovadores que guiam o processo de documentação digital, como a *Resource Description and Access (RDA)* e a ontologia PROV, que trazem meios para assegurar a integridade e a proveniência dos dados.

2.1 RDA

A RDA, desenvolvida inicialmente como uma evolução das Regras Anglo-Americanas de Catalogação, visa atender às demandas do ambiente digital, proporcionando maior flexibilidade e precisão na criação de registros bibliográficos. Baseada em modelos conceituais internacionais, a RDA incorpora princípios de controle de autoridade e acessibilidade, permitindo que metadados sejam gerados e utilizados de forma interoperável em diferentes contextos.

2.1.1 Histórico da RDA

A RDA foi inicialmente criada como a terceira edição das Regras Anglo-Americanas de Catalogação (AACR3), desenvolvida pelo Comitê Diretivo Conjunto para a Revisão do AACR, que seguiu seu Plano Estratégico de 2005-2009. Esse plano estratégico propunha mudanças estruturais, conceituais e terminológicas em relação ao AACR2, com o objetivo de alinhar o padrão às iniciativas internacionais para melhorar o acesso a recursos no ambiente digital (KINCY, 2014).

A RDA foi fundamentada em uma série de padrões e princípios de catalogação internacionais, destacando-se os Requisitos Funcionais para Registros Bibliográficos (FRBR),

um modelo conceitual introduzido em 1998, e seria compatível com outros padrões de metadados, enfatizando o conceito de controle de autoridade (KINCY, 2014).

O Plano Estratégico também pretendia fornecer diretrizes mais gerais para a catalogação de recursos de todos os tipos, adaptáveis a formatos emergentes ao longo do tempo. A Parte I do AACR2, que se restringia a certos tipos de formatos, impedia a inclusão de regras para novos formatos. Em contrapartida, o AACR3 seria reformulado para oferecer orientações específicas para diferentes tipos de conteúdo e mídia, eliminando redundâncias presentes nos capítulos baseados em formatos da Parte I do AACR2 (KINCY, 2014).

O AACR3 seria dividido em três partes principais: a Parte I contendo regras gerais e suplementares para tipos específicos de conteúdo e mídia; a Parte II baseada no Capítulo 21 do AACR2, focada em "Escolha de Pontos de Acesso" e oferecendo diretrizes para a inclusão de pontos de acesso que representam relações entre recursos e entre recursos e as entidades responsáveis por eles; e a Parte III abordaria os princípios de controle de autoridade, fornecendo orientações para a formulação de pontos de acesso autorizados e suas variantes (KINCY, 2014).

Com base no feedback recebido, o JSC (*Joint Steering Committee for Development of RDA*) ou Comitê Conjunto para o Desenvolvimento da RDA, órgão responsável por supervisionar o desenvolvimento inicial e as atualizações da RDA, decidiu adotar uma nova abordagem para a revisão do AACR, reconhecendo que os desafios apresentados pelo ambiente digital exigiam um padrão novo e mais adequado para a descrição e o acesso aos recursos, a RDA. Após a reunião de 2005, o JSC revisou e reorganizou o padrão, produzindo rascunhos de texto entre dezembro de 2005 e novembro de 2008 para avaliação. Em junho de 2009, o texto completo do novo padrão foi entregue aos editores da RDA e, em 2010, a RDA foi oficialmente lançado como parte de um produto online, o RDA Toolkit (KINCY, 2014).

A RDA representa um avanço significativo em relação ao AACR2, desenvolvido para atender de forma mais eficaz às demandas dos usuários finais e adaptar-se com maior facilidade às diversas comunidades. Este padrão de catalogação é embasado nos conceitos dos FRBR (Requisitos Funcionais para Registros Bibliográficos) e FRAD (Requisitos Funcionais para Dados de Autoridade), incorporando muitas das diretrizes do AACR2. Essa inclusão permite a coexistência harmoniosa de registros criados sob ambos os padrões. No entanto, a

RDA distingue-se do AACR2 por simplificar suas diretrizes e adotar uma terminologia mais congruente com outros padrões de metadados, evitando termos obscuros para facilitar a compreensão e o uso dos dados pelos usuários contemporâneos (KINCY, 2014).

Em contraste com o AACR3, que foi inicialmente pensado como um produto físico, a RDA foi projetada para operar predominantemente em um ambiente digital. Como resultado, o principal meio de acesso a RDA é o RDA Toolkit, um recurso online que é complementado por versões impressas publicadas de forma periódica. A estrutura organizacional da RDA também sofreu uma transformação radical: o Comitê Diretivo Conjunto (JSC) abandonou o formato de três partes do AACR2, o qual optou por dividir a RDA em trinta e sete capítulos distribuídos em dez seções, todas baseadas nos princípios dos FRBR e FRAD (KINCY, 2014).

Em 2017, o Projeto 3R foi implementado, trazendo uma reformulação a RDA. Esse projeto, descrito por Oliver (2021), modernizou e reestruturou o padrão para adaptá-lo às demandas do ambiente digital, com melhorias substanciais no RDA Toolkit, como uma interface mais amigável e ferramentas de edição aprimoradas.

A transição para a RDA Oficial se deu devido à necessidade de um padrão de catalogação que esteja alinhado aos modelos internacionais e seja capaz de lidar com a crescente complexidade dos ambientes digitais. A RDA foi projetada para ser flexível e extensível, permitindo a incorporação de novos formatos e tipos de conteúdo de maneira eficiente, algo que não era possível com padrões anteriores como o AACR2. Além disso, a compatibilidade com tecnologias emergentes, como a Web Semântica, garante que os metadados sejam interoperáveis em diferentes contextos e instituições, promovendo uma maior transparência e acessibilidade das informações (OLIVER, 2021).

Outro aspecto relevante da RDA é sua capacidade de integrar-se a outras ontologias e padrões, como a PROV-O, o que facilita o rastreamento da proveniência dos dados em ambientes bibliográficos e culturais. Isso é importante para garantir a integridade e a autenticidade das informações, especialmente em um cenário digital globalizado. A RDA também moderniza o processo de catalogação, separando as instruções de registro de dados das de apresentação, permitindo maior flexibilidade na maneira como as informações são armazenadas e exibidas (KINCY, 2014).

Essas mudanças foram fundamentais para garantir que a RDA continuasse a atender às necessidades de catalogação e descrição de recursos em um contexto global, facilitando sua adoção em diferentes realidades culturais e tecnológicas, ao mesmo tempo em que estabeleceu diretrizes para uma ampla variedade de formatos, com atenção especial aos digitais, introduzindo a gravação de dados de autoridade, uma melhora considerável em relação ao AACR2 (OLIVER, 2021).

Portanto, a RDA constitui uma reformulação substancial do AACR2, apoiada nos mesmos princípios fundamentais de catalogação, mas expandida para incluir novas metas e princípios destinados a contextualizar tanto as mudanças quanto as continuidades entre os dois padrões. Essa abordagem mais moderna busca não apenas preservar a consistência com a tradição da catalogação, mas também ampliar sua aplicabilidade e usabilidade em contextos digitais (KINCY, 2014).

2.1.2 Estrutura da RDA

A RDA tem como um dos principais objetivos atender às necessidades dos usuários finais, garantindo que os dados que descrevem os recursos sejam úteis para uma variedade de tarefas de busca e uso de informação. De acordo com Kincy (2014), a RDA orienta os catalogadores a registrarem atributos e relações que facilitam a localização de recursos relevantes, como obras, pessoas, famílias ou entidades corporativas e seus vínculos com outros recursos ou entidades.

O novo código da RDA reflete o fluxo de trabalho esperado para a descrição de recursos, começando com as regras para elaborar a representação de manifestação e item, seguidas das normas para obra e expressão, e terminando com as instruções para representar as relações. A estrutura da RDA facilita o processo de catalogação, adaptando-se melhor aos avanços digitais (ALMIRÓN et al., 2015).

Adicionalmente, a RDA auxilia os usuários na seleção de recursos com base em características como forma física, público-alvo e idioma, através de capítulos específicos que tratam das características de suporte e conteúdo dos recursos. O padrão também oferece diretrizes para facilitar o acesso aos recursos, seja por compra, empréstimo ou acesso online. A RDA também destaca a importância de entender as relações entre entidades, as diferentes formas dos nomes e a lógica por trás da escolha de formas autorizadas e variantes. Para isso,

são fornecidas instruções específicas sobre controle de autoridade, que incluem o registro de informações que esclarecem essas relações e escolhas, contribuindo para a compreensão e utilização eficaz dos dados pelos usuários (KINCY, 2014).

A implementação da RDA no cenário da Biblioteconomia brasileira apresenta desafios e oportunidades. De acordo com Mey, Grau e Biar (2014), a RDA, ao propor-se como um código internacional, incorpora inovações importantes, como a inclusão de materiais digitais e elementos do modelo *Functional Requirements for Bibliographic Records* (FRBR). Essas características ampliam as possibilidades de descrição de recursos, mas também impõem a necessidade de adaptação aos contextos locais, especialmente em termos de tradução e licenciamento.

A tradução do RDA para o português representa um ponto de atenção, pois as limitações linguísticas e os custos envolvidos podem restringir sua aplicação ampla. Conforme destacado, “o problema das restrições na tradução e seus custos de licenciamento e direitos autorais” é um dos fatores que dificultam a adoção plena da RDA no Brasil (MEY; GRAU; BIAR, 2014). Essa barreira evidencia a importância de desenvolver estratégias que equilibrem a padronização internacional com as necessidades locais de cada instituição.

Também foram introduzidas mudanças na forma como os pontos de acesso são tratados, particularmente em relação a nomes de pessoas, famílias e conferências. Essas mudanças oferecem uma maior clareza e consistência ao registrar a forma preferida dos nomes e informações associadas, facilitando a recuperação de dados pelos usuários (ALMIRÓN et al., 2015).

Um dos objetivos da RDA é a eficiência de custos, a qual busca que os dados atendam às necessidades funcionais de suporte às tarefas do usuário de maneira eficiente, conforme Kincy (2014). O processo é simplificado ao eliminar várias das convenções de catalogação prescritas pelo AACR2, tais como regras sobre abreviações e omissões.

A flexibilidade é outro ponto no qual os dados devem ser funcionais independente do formato, meio ou sistema usado para armazená-los ou comunicá-los. A RDA foi desenvolvida para ser adequada em uma variedade de meios, indo além das limitações do catálogo de fichas tradicional do AACR2. As instruções não prescrevem a truncagem ou omissão de dados baseadas em convenções de economia de espaço e evitam a terminologia específica dos

catálogos de fichas. Além disso, são escritas de forma a considerar atributos e relações, aplicáveis tanto em contextos MARC ou não (KINCY, 2014).

Por conseguinte, a continuidade é um aspecto notável da RDA, que busca a integração adequada dos dados em bases de dados existentes, muitas das quais foram desenvolvidas usando padrões do AACR. Apesar das inovações, muitas das regras do AACR2 foram incorporadas a RDA com modificações mínimas. Isso facilita a coexistência do AACR2 e da RDA dentro das mesmas bases de dados, o que permite uma transição mais fácil e com menor dificuldade (KINCY, 2014).

Além dos aspectos técnicos e estruturais, é importante considerar as implicações culturais e práticas que a transição do AACR2 para a RDA trouxe para a comunidade bibliotecária. Conforme apontado por Modesto (2013), a introdução da RDA não apenas modernizou os processos de catalogação, mas também estimulou debates sobre o papel do catalogador e as novas exigências do ambiente digital. Nesse contexto, Modesto (2013) compara a transição a RDA a uma transformação religiosa, destacando o caráter 'profético' do novo padrão e seu potencial para remodelar as práticas de catalogação.

Modesto (2013) também argumenta que a RDA é uma norma 'agnóstica' que transcende formatos tradicionais, permitindo a descrição de uma ampla gama de recursos, sejam físicos ou digitais. Essa flexibilidade, segundo o autor, representa uma mudança de paradigma, que facilita a adaptação dos catalogadores a novas tecnologias e métodos. Além disso, o autor destaca o impacto econômico e cultural da implementação da RDA, mencionando as dificuldades financeiras que as bibliotecas brasileiras podem enfrentar ao adquirir licenças para o RDA Toolkit. Modesto critica a comercialização das normas de catalogação, sugerindo que essa prática pode limitar o acesso a ferramentas essenciais para a organização da informação.

A modernização proporcionada pela RDA abrange tanto bibliotecas, como museus e arquivos, promovendo uma maior interoperabilidade entre diferentes sistemas de informação. Essa característica da RDA possibilita a adoção de um padrão comum para a descrição e acesso a dados de diversas instituições culturais (MODESTO, 2013).

A *Resource Description and Access* (RDA) é um padrão internacional de catalogação desenvolvida para fornecer uma estrutura abrangente e consistente de descrição e acesso aos

recursos informacionais em ambientes digitais e tradicionais. Esse padrão foi projetado para ser usado além dos registros bibliográficos convencionais, abrangendo também dados de autoridades e informações de proveniência, permitindo maior interoperabilidade e reutilização de dados em diferentes contextos bibliográficos (OLIVER, 2023; RDA TOOLKIT, 2022).

Com sua base no modelo conceitual do FRBR e seus sucessores, a RDA adota uma abordagem centrada nas entidades, promovendo uma representação rica e detalhada dos recursos bibliográficos e suas relações (KINCY, 2014).

No contexto da RDA, as entidades são fundamentais para a estruturação dos dados bibliográficos e de autoridade. O modelo RDA foi idealizado em torno do conceito de entidades, que constituem os elementos essenciais para representar informações bibliográficas. Essas entidades são estruturadas conforme o modelo FRBR e seus sucessores, abrangendo uma ampla gama de elementos que descrevem as relações entre recursos, seus criadores e seus assuntos (RDA TOOLKIT, 2022).

A entidade principal, conhecida como RDA Entity, é uma classe abstrata que representa objetos conceituais no universo do discurso humano, os quais são de interesse para os usuários de metadados RDA. Dentro dessa classe geral, encontra-se várias entidades específicas, começando com a Obra (*Work*), que é definida como uma criação intelectual ou artística distinta. É a base para todas as outras entidades que derivam dela, como a Expressão (*Expression*), que é a realização concreta de uma obra em uma forma particular, seja através de uma tradução, adaptação ou outra forma de expressão artística (RDA TOOLKIT, 2022).

Seguindo a expressão, a Manifestação (*Manifestation*) é a concretização física ou digital dessa expressão. Essa entidade refere-se a aspectos como a forma de publicação, se o formato é físico ou digital, e outros detalhes que identificam como uma expressão é apresentada ao público. Já o Item é uma instância específica de uma manifestação, caracterizada por suas peculiaridades, como anotações, marcas de propriedade, ou outros sinais que tornam um item único dentro de uma coleção (RDA TOOLKIT, 2022).

No campo das entidades relacionadas a agentes, o Agente (*Agent*) é uma categoria extensa que inclui indivíduos e grupos responsáveis por ações deliberadas associadas a obras, expressões, manifestações e itens. Dentro dessa entidade, estão: Pessoa (*Person*), que se

refere a um indivíduo humano; Família (*Family*), que abrange grupos de pessoas relacionadas por laços de parentesco; e a Entidade Corporativa (*Corporate Body*), que representa organizações ou grupos de pessoas organizadas para atingir um objetivo comum (RDA TOOLKIT, 2022).

Outras entidades no modelo RDA incluem Lugar (*Place*), que se refere a uma localização geográfica específica associada a uma entidade ou a um recurso; Período de Tempo (*Timespan*), que descreve um intervalo de tempo definido, relevante para a catalogação e organização dos recursos. E finalmente, *Nomen*, que é uma entidade que representa rótulos ou identificadores que podem ser aplicados a qualquer outra entidade RDA, como nomes, títulos, ou outros identificadores usados para descrever ou acessar um recurso (RDA TOOLKIT, 2022).

Para assegurar a conformidade dos registros bibliográficos com os padrões de interoperabilidade e integridade de dados, é fundamental examinar como os elementos da RDA se alinham com os princípios de proveniência, como definidos pelo modelo PROV-O. A proveniência, no contexto da catalogação e gestão de dados, refere-se à documentação detalhada do histórico de criação, uso e transformação de um recurso ao longo do tempo. Este processo é essencial para a validação e confiabilidade dos dados em sistemas bibliográficos, garantindo que cada recurso possa ser rastreado até sua origem e que suas modificações ao longo do tempo sejam claramente compreendidas (W3C, 2013). Considerando a importância desse processo, a análise dos elementos selecionados do RDA Toolkit é um importante passo para verificar a aderência aos padrões de proveniência estabelecidos pelo PROV-O (OLIVER, 2021).

Com isso, após análise e identificação dos principais elementos da RDA, foram selecionados alguns elementos do RDA Toolkit para uma análise detalhada quanto ao grau de aderência aos recursos de proveniência definidos pelo PROV-O. Esses elementos, totalizando dezessete (17), juntamente com suas descrições, estão apresentados na tabela a seguir, onde serão examinados posteriormente quanto à sua compatibilidade com os princípios de proveniência.

Tabela 1 - Elementos selecionados do RDA Toolkit.

Elementos	Descrição
<i>rda:RDA Entity</i> (Entidade RDA)	É uma classe abstrata de objetos conceituais chave no universo do discurso humano, que são de interesse para os usuários dos metadados RDA em um sistema de descoberta de recursos. As entidades RDA incluem agentes, corpos coletivos, entidades corporativas, expressões, famílias, itens, manifestações, <i>nomens</i> , pessoas, lugares, períodos de tempo e obras.
<i>rda:Expression</i> (Expressão)	Entidade que é definida pela realização específica de uma obra, que pode ser manifestada de várias formas. Por exemplo, uma obra literária pode ter várias expressões, como traduções, edições revisadas ou versões adaptadas. Cada uma representa uma variação específica do conteúdo intelectual ou artístico de uma obra.
<i>rda:Agent</i> (Agente)	Entidade capaz de realizar ações deliberadas, ser responsabilizada por elas, e inclui pessoas, famílias e corporações.
<i>rda:Generation</i> (Geração)	Representa o processo ou evento que resulta na criação de uma entidade específica, como um trabalho ou expressão
<i>rda:derivative work relationship</i> (relação de obra derivada)	Refere-se à relação entre uma obra derivada e a obra original da qual ela foi adaptada ou modificada. Esse tipo de relação descreve como uma nova obra foi criada a partir de uma existente, seja por meio de adaptação, tradução etc.
<i>rda:statement of responsibility</i> (declaração de responsabilidade)	Refere-se a uma declaração que identifica ou descreve a função de qualquer pessoa, família ou corpo corporativo responsável pela criação, ou que contribuiu para a realização, do conteúdo intelectual ou artístico de um recurso.
<i>rda:beginning of</i> (início de)	Descreve o início de um período de tempo, utilizado para indicar quando um intervalo de tempo ou uma condição específica começa.
<i>rda:source consulted</i> (fonte consultada)	Refere-se a uma fonte consultada durante a determinação de um nome, título ou outro atributo identificador de uma entidade, ou durante a determinação da relação entre entidades. É utilizado para citar a(s) fonte(s) de onde as informações foram obtidas, seguindo de uma breve declaração sobre a informação encontrada.
<i>rda:ending of</i> (fim de)	Relaciona um período de tempo a um ponto final específico. É utilizado para indicar o término de um evento,

	atividade ou condição dentro de um determinado período de tempo.
<i>rda:related agent of agent</i> (agente relacionado do agente)	Relaciona um agente a outro agente associado. Usado para descrever uma associação entre agentes no contexto de um recurso ou atividade.
<i>rda:related collective agent of agent</i> (agente coletivo relacionado do agente)	Relaciona um agente coletivo a outro agente associado. Usado para descrever a associação entre agentes coletivos e outras entidades ou agentes.
<i>rda:related work of agent</i> (obra relacionada ao agente)	Descreve a relação entre uma obra e o agente responsável por ela, especialmente quando há múltiplas versões ou edições atribuídas ao mesmo agente.
<i>rda:related expression of work</i> (expressão relacionada à obra)	Descreve a relação entre uma obra e suas diferentes expressões, como traduções, adaptações, ou edições específicas.
<i>rda:date of production</i> (data de produção)	Representa a data em que uma manifestação foi fisicamente produzida. Refere-se ao momento em que o recurso foi materialmente criado ou fabricado.
<i>rda:source consulted</i> (fonte consultada)	Refere-se à fonte ou recurso utilizado como base para a criação de uma obra ou expressão. Pode incluir referências a materiais que influenciaram diretamente a criação do recurso.
<i>rda:revision of</i> (revisão de)	Representa uma relação entre uma obra ou expressão e suas revisões, indicando que a obra atual é uma versão revisada de uma anterior.
<i>rda:location</i> (localização)	Refere-se a um lugar específico ou posição geográfica associada a uma entidade ou evento, usado para indicar onde algo está localizado ou onde ocorre.

Fonte: RDA Toolkit (2024).

Após a análise detalhada dos elementos da RDA, é essencial compreender como a proveniência de dados é estruturada em um contexto mais amplo. O modelo PROV, desenvolvido pelo *World Wide Web Consortium (W3C)*, oferece uma estrutura confiável para representar a proveniência de dados em ambientes digitais.

Na próxima seção, serão explorados os conceitos fundamentais do PROV e como eles podem ser aplicados à gestão e rastreamento de dados, a fim de garantir a transparência, rastreabilidade e autenticidade dos recursos descritos.

2.2 PROV

A ontologia PROV, estabelecida pelo *World Wide Web Consortium* (W3C), foi criada para fornecer um modelo robusto e padronizado de proveniência de dados em ambientes digitais. Estruturada com base em uma arquitetura semântica, o PROV-O permite a representação de informações sobre a origem, transformação e interações de dados, garantindo rastreabilidade e transparência. Sua flexibilidade e compatibilidade com outros padrões a tornam uma ferramenta para gerenciar e documentar fluxos de dados em diversas aplicações, assegurando a interoperabilidade entre sistemas.

2.2.1 Conceitos de proveniência

No contexto do modelo PROV, desenvolvido pelo W3C, a proveniência de dados é tratada de forma detalhada e estruturada, proporcionando uma representação abrangente das interações e do histórico dos dados em ambientes digitais.

Proveniência é entendida como o registro das entidades, atividades e agentes envolvidos na produção ou modificação de um recurso, o que permite a rastreabilidade da origem e das transformações sofridas por determinado dado. Tal definição é particularmente relevante para contextos em que a integridade e a confiabilidade das informações são primordiais, alinhando-se diretamente às necessidades de sistemas de metadados, como o RDA (MOREAU et al., 2015).

Os fundamentos do modelo PROV organizam a proveniência em três elementos principais: entidades, atividades e agentes. Essa estrutura possibilita a representação formal de informações de proveniência, promovendo a interoperabilidade entre diferentes sistemas. No contexto do RDA, que também adota uma abordagem estruturada para descrever recursos e suas relações, o PROV oferece uma base complementar, destacando a importância de um modelo padronizado para garantir consistência e integração de dados em sistemas bibliográficos (MOREAU et al., 2015).

As atividades são processos ou eventos que utilizam, transformam ou geram entidades, ocorrendo em um intervalo de tempo específico. Essas atividades são fundamentais para entender como os dados são modificados ou produzidos, ajudando a mapear o ciclo de vida dos dados. Já os agentes são indivíduos, organizações ou sistemas responsáveis por entidades ou atividades. Eles possuem a capacidade de iniciar, controlar ou

influenciar atividades, afetando diretamente a criação ou modificação de entidades (ARAKAKI, 2020).

Além desses elementos básicos, o modelo PROV introduz conceitos como responsabilidade e derivação. A responsabilidade reflete a atribuição de agentes a atividades específicas, enquanto a derivação documenta como entidades são originadas ou alteradas ao longo do tempo, reforçando a transparência no rastreamento de dados (ARAKAKI, 2020; W3C, 2013).

O modelo PROV também suporta o uso de pacotes de informações ou *bundles*, que são conjuntos de dados de proveniência organizados para documentar a proveniência de outros dados, criando uma camada transparente para auditoria e rastreamento dos dados (ARAKAKI, 2020).

O desenvolvimento do modelo PROV foi fundamentado em princípios como simplicidade, generalidade e usabilidade. Esses princípios foram concebidos para garantir que o modelo fosse amplamente aplicável, ao mesmo tempo em que assegurasse sua funcionalidade em contextos variados. A simplicidade visa reduzir a complexidade no uso do modelo, enquanto a generalidade permite sua aplicação em diferentes domínios. Já a usabilidade assegura que o modelo seja acessível tanto para humanos quanto para máquinas, promovendo sua adoção em ambientes diversos, como bibliotecas e repositórios digitais (MOREAU et al., 2015).

Outro aspecto relevante do modelo PROV é sua capacidade de permitir a validação e verificação de dados de proveniência. O modelo provê mecanismos para assegurar que os registros de proveniência sejam consistentes e possam ser auditados, o que contribui para a confiabilidade das informações. Essa funcionalidade é crucial em contextos em que a rastreabilidade e a integridade dos dados são requisitos essenciais, como em acervos digitais e sistemas de catalogação bibliográfica (MOREAU et al., 2015).

O PROV também se destaca por sua extensibilidade, permitindo que novos conceitos e relações sejam adicionados conforme as necessidades dos usuários e dos sistemas evoluem. Essa característica garante que o modelo possa ser adaptado a diferentes domínios e requisitos específicos, como aqueles encontrados em bibliotecas e arquivos digitais. A possibilidade de customização do modelo reforça sua aplicabilidade e longevidade em

sistemas que necessitam de soluções robustas e flexíveis para a gestão de metadados (MOREAU et al., 2015).

Esses conceitos são fundamentais para a implementação prática do modelo PROV em ambientes digitais, oferecendo uma base sólida para a gestão de metadados e garantindo a integridade e confiabilidade dos dados, fatores essenciais em domínios como a ciência de dados, a biblioteconomia e a de acervos digitais (ARAKAKI, 2020).

2.2.2 Família PROV

De acordo com Arakaki (2020), o modelo PROV foi desenvolvido para facilitar a interoperabilidade e a extensibilidade dos dados de proveniência entre diferentes sistemas. Criada pelo W3C, a família PROV foi projetada para oferecer uma representação sólida de proveniência em contextos digitais. Esses documentos oferecem uma abordagem padronizada para documentar a origem e a história dos dados digitais, contribuindo assim para a confiabilidade, reutilização e compreensão dos mesmos.

As quatro recomendações da família PROV, segundo Arakaki (2020), são:

- PROV-DM (*Data Model*): é o modelo de dados conceitual fundamental, que define os tipos de objetos e as relações entre eles no contexto da proveniência. Inclui entidades, atividades e agentes, com seus respectivos atributos e interações.
- PROV-N (*Notation*): é uma notação legível tanto para máquinas quanto para seres humanos, destinada a descrever exemplos e instâncias de proveniência de acordo com o PROV-DM. Ela facilita a expressão e a troca de informações de proveniência entre diferentes sistemas.
- PROV-O (*Ontology*): é uma ontologia baseada na *Web Ontology Language* (OWL), que permite a incorporação de dados de proveniência em sistemas de dados da web. O PROV-O facilita a integração de informações de proveniência em ambientes semânticos da web, oferecendo uma série de classes e propriedades projetadas para mapear diretamente para os elementos definidos no PROV-DM.
- PROV-CONSTRAINTS: define restrições sobre instâncias de proveniência para garantir a validade dos dados registrados sob o modelo PROV. Essas restrições

ajudam a assegurar que as informações de proveniência são consistentes e lógicas ao longo de diferentes registros e aplicações.

No desenvolvimento do modelo PROV, a intenção do W3C foi criar um sistema que atendesse às necessidades iniciais dos usuários e desenvolvedores para questões básicas da proveniência, e que também permitisse a evolução para aplicações mais complexas conforme a evolução crescente da experiência e das necessidades dos usuários (ARAKAKI, 2020).

2.2.3 Estrutura da Ontologia PROV (PROV-O)

Na concepção da ontologia PROV, o PROV-O, o W3C estabelece uma infraestrutura sólida para a representação interoperável de metadados de proveniência em ambientes digitais. De acordo com Arakaki (2020), PROV-O é baseada na *Web Ontology Language* (OWL) e é essencial para que a proveniência de forma que possa ser entendida e utilizada por diferentes sistemas, promovendo uma vasta gama de aplicações, desde pesquisas científicas até a gestão governamental de dados.

A ontologia organiza-se em torno de três classes principais (ARAKAKI, 2020):

- Entidades (*prov:Entity*): Objetos físicos, digitais ou conceituais que possuem identidade fixa. Essas entidades podem ser tanto concretas quanto abstratas, e a ontologia permite detalhar suas características e transformações ao longo do tempo.
- Atividades (*prov:Activity*): Processos ou ações que modificam, influenciam ou geram entidades. Essas atividades são descritas em termos de seu desenvolvimento no tempo, interações com entidades e influências de agentes.
- Agentes (*prov:Agent*): Representações de indivíduos, organizações ou sistemas que desempenham um papel ativo nas atividades, sendo responsáveis por mudanças ou pela geração de novas entidades.

Além desses elementos, o PROV-O facilita a descrição de relações complexas entre entidades, atividades e agentes, incluindo a geração de entidades por atividades, a influência de agentes sobre as atividades, e as múltiplas camadas de responsabilidade e autoria (Arakaki, 2020).

Muitos são os benefícios de implementar o PROV-O, conforme Arakaki (2020) destaca que a capacidade de detalhar a proveniência não só aumenta a transparência e a confiabilidade, mas também fortalece a integridade dos dados em ambientes que podem demandar alto nível de precisão e auditoria.

A ontologia PROV é uma estrutura para a gestão de dados digitais, permitindo que profissionais de diversas áreas capturem, descrevam e compartilhem a proveniência de informações de maneira eficaz e normatizada. Isso é essencial para garantir a confiança e a verificação em processos que dependem de dados precisos e bem documentados, além de facilitar o reuso dos dados, uma vez que estes são considerados confiáveis (ARAKAKI, 2020).

O PROV-O é dividido em três categorias principais de termos: *Starting Point Terms*, *Expanded Terms* e *Qualified Terms*. Esses termos fornecem a base para a ontologia PROV, permitindo descrições simples de proveniência (W3C, 2013).

Os *Starting Point Terms* do PROV-O são um conjunto de termos fundamentais que servem como base para descrever a proveniência de dados, atividades e agentes. Eles são projetados para serem intuitivos e fornecer uma maneira simples para as informações de proveniência. Esses termos são essenciais para estabelecer uma compreensão inicial das relações e interações que ocorrem entre entidades, atividades e agentes (W3C, 2013).

Nos *Starting Point Terms* do PROV-O, encontramos as seguintes três (3) classes e nove (9) propriedades na tabela abaixo.

Tabela 2 - *Starting Point Terms* do PROV-O.

	Elementos	Descrição
Classes	<i>prov:Entity</i> (Entidade)	Representa qualquer coisa física, digital, conceitual ou de outro tipo, com aspectos fixos; pode ser real ou imaginária.
	<i>prov:Activity</i> (Atividade)	Algo que ocorre ao longo de um tempo e pode incluir processos e ações que geram ou utilizam entidades.
	<i>prov:Agent</i> (Agente)	Indica que uma entidade foi atribuída a um agente, sugerindo responsabilidade ou associação.
Propriedades	<i>prov:wasGeneratedBy</i> (foi Gerado Por)	Indica que uma entidade foi gerada por uma atividade.

<i>prov:wasDerivedFrom</i> (foi Derivado De)	Indica que uma entidade A foi derivada de outra entidade B. A relação de derivação implica que A resulta de B, possivelmente através de uma transformação ou modificação.
<i>prov:wasAttributedTo</i> (foi Atribuído a)	Indica que uma entidade foi atribuída a um agente, sugerindo responsabilidade ou associação.
<i>prov:startedAtTime</i> (começou No Tempo)	Indica o momento em que uma atividade começou. A atividade não existia antes desse momento. Pode estar relacionado a uma entidade que desencadeou a atividade ou a uma atividade que gerou essa entidade.
<i>prov:used</i> (Usado)	Representa o início do uso de uma entidade por uma atividade. Antes de seu uso, a atividade não poderia ter sido afetada por essa entidade.
<i>prov:wasInformedBy</i> (foi Informado Por)	Indica que uma atividade foi influenciada por outra atividade. Especifica que a informação de uma atividade serviu de base ou influenciou outra atividade, descrevendo uma relação de dependência informacional.
<i>prov:endedAtTime</i> (finalizado No Tempo)	Representa o momento em que uma atividade é considerada encerrada. Este elemento marca o fim de uma atividade, e qualquer uso, geração ou invalidação envolvendo uma atividade precede seu término.
<i>prov:wasAssociatedWith</i> (foi Associado a)	Associa uma atividade a um agente, indicando que o agente foi responsável por realizar a atividade.
<i>prov:actedOnBehalfOf</i> (agiu Em Nome De)	Indica que um agente atuou em nome de outro agente, delegando responsabilidade.

Fonte: W3C (2013).

Os *Expanded Terms* do PROV-O são uma extensão dos *Starting Point Terms* e fornecem de forma mais detalhada a descrição das relações de proveniência entre entidades, atividades

e agentes. Apresentam termos adicionais para casos mais complexos e específicos. Abaixo segue uma tabela com esses termos divididos em Classes e Propriedades (W3C, 2013).

Nos *Expanded Terms* do PROV-O, encontramos as seguintes sete (7) classes e dezesseis (16) propriedades na tabela abaixo.

Tabela 3 - *Expanded Terms* do PROV-O.

	Elementos	Descrição
Classes	<i>prov:Collection</i> (Coleção)	Conjunto de entidades agrupadas com base em uma característica comum.
	<i>prov:EmptyCollection</i> (Coleção Vazia)	Representa uma coleção que, por design ou intenção, não contém nenhuma entidade.
	<i>prov:Bundle</i> (Pacote)	Entidade que agrupa e identifica de forma única um conjunto de declarações de proveniência. Permite a representação e o gerenciamento coletivo de várias entidades, atividades e agentes, facilitando a organização e a relação dessas informações de proveniência em um único contexto.
	<i>prov:Person</i> (Pessoa)	Um agente que é um ser humano. Representa pessoas que atuam ou são referenciadas em contextos de proveniência, desempenhando papéis específicos em atividades e eventos.
	<i>prov:SoftwareAgent</i> (Agente de Software)	Um agente que é um programa de software, ou seja, uma aplicação ou sistema que executa ações de forma automatizada ou sem necessidade de intervenção humana direta.
	<i>prov:Organization</i> (Organização)	Um agente que é uma instituição ou grupo de pessoas organizadas com um propósito comum, como empresas, organizações governamentais, ONGs etc.
	<i>prov:Location</i> (Localização)	Refere-se a uma posição ou lugar físico onde algo ocorre ou está situado. Pode incluir localizações geográficas ou outras áreas definidas no contexto de proveniência.

Propriedades	<i>prov:alternateOf</i> (alternativo De)	Indica que duas entidades representam a mesma coisa em diferentes contextos ou que uma é uma alternativa à outra.
	<i>prov:specializationOf</i> (especialização De)	Indica que uma entidade é uma forma mais específica de outra entidade. Representa uma relação hierárquica, onde uma entidade especializada mantém uma conexão com uma entidade mais geral.
	<i>prov:generatedAtTime</i> (gerado Em)	Indica o momento exato ou período de tempo em que uma entidade foi gerada ou criada por uma atividade.
	<i>prov:hadPrimarySource</i> (Fonte Primária envolvida)	Indica que uma entidade teve uma fonte primária, ou seja, uma entidade anterior que serviu como a principal base de evidência ou informação.
	<i>prov:value</i> (valor)	Indica o valor de uma entidade, geralmente associado a dados ou informações concretas.
	<i>prov:wasQuotedFrom</i> (foi Citado De)	Indica que uma entidade contém conteúdo que foi extraído de outra entidade, como uma citação direta de um texto ou outra forma de conteúdo.
	<i>prov:wasRevisionOf</i> (foi Revisão De)	Indica que uma entidade é uma versão revisada de outra entidade, sugerindo uma modificação ou atualização em relação à versão anterior.
	<i>prov:invalidatedAtTime</i> (invalidado no Tempo)	Indica o momento exato ou período de tempo em que uma entidade foi invalidada, ou seja, deixou de ser válida, utilizável ou existente.
	<i>prov:wasInvalidatedBy</i> (foi Invalidado Por)	Indica que uma entidade foi invalidada ou que deixou de existir por uma atividade específica.
	<i>prov:hadMember</i> (Membros envolvidos)	Indica a relação de pertencimento entre uma coleção e seus membros. Estabelece uma ligação entre uma entidade que é um membro e a coleção à qual ela pertence.
<i>prov:wasStartedBy</i> (foi Iniciado Por)	Indica que uma atividade foi iniciada por uma entidade específica. Estabelece uma relação causal direta, mostrando que a	

		entidade mencionada desempenhou um papel fundamental no início da atividade.
	<i>prov:wasEndedBy</i> (foi Finalizado Por)	Indica que uma atividade foi finalizada por uma entidade específica. Estabelece uma relação causal direta, mostrando que a entidade mencionada desempenhou um papel importante na finalização da atividade.
	<i>prov:invalidated</i> (invalidado)	Indica que uma entidade foi invalidada ou deixou de existir como resultado de uma atividade específica. É usada para registrar o momento em que uma entidade perdeu sua validade ou foi retirada de circulação, associando esse evento a uma atividade que causou a invalidação.
	<i>prov:influenced</i> (influenciado)	Indica que uma entidade, atividade ou agente foi influenciado por outra entidade, atividade ou agente. É usado para registrar a ideia de que um elemento teve um impacto ou efeito sobre outro, sem especificar a natureza exata dessa influência.
	<i>prov:atLocation</i> (no Local)	Refere-se a um lugar ou localização onde algo ocorre ou está situado. Pode ser um lugar geográfico ou uma localização não-geográfica, como um diretório ou uma célula em uma tabela.
	<i>prov:generated</i> (gerado)	Indica que uma entidade foi gerada como resultado de uma atividade, marcando a criação ou produção de algo específico.

Fonte: W3C (2013).

Já as *Qualified Classes and Properties* oferecem informações detalhadas sobre as relações binárias, estabelecidas por meio das propriedades dos *Starting Point Terms* e dos *Expanded Terms*. Ao contrário das categorias anteriores, onde as relações são expressas de maneira direta e binária, nesta categoria, os termos são utilizados para atribuir características adicionais às relações binárias (W3C, 2013).

Essa abordagem permite aos usuários descrições mais detalhadas que não poderiam ser descritos somente com os *Starting Point Terms* e os *Expanded Terms*. Assim, as *Qualified*

Classes and Properties são fundamentais para uma descrição mais precisa e completa das relações, já que permite que sejam fornecidos detalhes que não estão disponíveis apenas usando os *Starting Point Terms* e os *Expanded Terms* (W3C, 2013).

Nas *Qualified Classes and Properties* do PROV-O, encontramos as seguintes vinte (20) classes e vinte e cinco (25) propriedades na tabela abaixo.

Tabela 4 - *Qualified Classes and Properties* do PROV-O.

	Elementos	Descrição
Classes	<i>prov:Influence</i> (Influência)	Representa a ideia de que uma entidade, atividade ou agente influenciou outro. Denota que o influenciador teve algum impacto ou efeito sobre o influenciado.
	<i>prov:EntityInfluence</i> (Influência de Entidade)	Representa a influência que uma entidade teve sobre outra, indicando que uma entidade influenciou o uso, criação, modificação ou existência de outra entidade ou atividade.
	<i>prov:Usage</i> (Uso)	Representa uma situação em que uma atividade utiliza uma entidade. Indica que uma entidade específica foi usada em uma atividade durante seu curso ou realização.
	<i>prov:Start</i> (Início)	Representa o momento em que uma atividade teve início. Indica a relação entre uma atividade e a entidade ou agente que marcou o começo dessa atividade.
	<i>prov:End</i> (Fim)	Representa o momento em que uma atividade terminou. Indica a relação entre uma atividade e a entidade ou agente que marcou o fim dessa atividade.
	<i>prov:Derivation</i> (Derivação)	Representa a relação em que uma entidade foi derivada de outra. Indica que uma entidade existe em virtude de modificações, transformações ou derivações de outra entidade.
	<i>prov:PrimarySource</i> (Fonte Primária)	Representa uma relação específica de derivação em que uma entidade foi

		derivada de outra entidade considerada uma fonte primária. Indica que a entidade derivada mantém uma forte relação com a entidade original.
	<i>prov:Quotation</i> (Citação)	Representa uma relação de derivação em que uma entidade foi derivada de outra por meio da citação direta. Indica que uma parte da entidade derivada é uma cópia exata ou uma reprodução da entidade original.
	<i>prov:Revision</i> (Revisão)	Representa uma relação de derivação em que uma entidade foi derivada de outra por meio de revisões. Indica que a entidade derivada é uma versão modificada da entidade original.
	<i>prov:ActivityInfluence</i> (Influência de Atividade)	Representa a influência que uma atividade teve sobre uma entidade, outra atividade ou agente. Indica que uma atividade afetou, direta ou indiretamente, o resultado ou a execução de outra atividade ou a criação de uma entidade.
	<i>prov:Generation</i> (Geração)	Representa o momento em que uma entidade foi gerada por uma atividade. Indica que uma entidade começou a existir como resultado de uma atividade específica.
	<i>prov:Communication</i> (Comunicação)	Representa a relação em que uma atividade se comunica com outra, indicando que uma atividade enviou ou compartilhou informações com outra atividade.
	<i>prov:Invalidation</i> (Invalidação)	Representa o momento em que uma entidade foi invalidada, indicando que uma entidade deixou de ser válida ou relevante como resultado de uma atividade.
	<i>prov:AgentInfluence</i> (Influência de Agente)	Representa a influência que um agente (pessoa, organização, software, etc.) teve sobre uma entidade, atividade ou outro agente. Indica que as ações ou decisões

		de um agente afetaram diretamente o resultado de uma atividade ou a criação/modificação de uma entidade.
	<i>prov:Attribution</i> (Atribuição)	Representa a relação entre uma entidade e um agente, indicando que um agente é responsável pela criação, modificação ou existência dessa entidade.
	<i>prov:Association</i> (Associação)	Representa a relação entre uma atividade e um agente, indicando que o agente teve um papel associado à realização dessa atividade.
	<i>prov:Plan</i> (Plano)	Representa um plano ou conjunto de intenções, procedimentos ou passos que orientam uma atividade, indicando que uma atividade foi realizada de acordo com um plano específico.
	<i>prov:Delegation</i> (Delegação)	Representa a relação em que um agente delega responsabilidade a outro agente para a realização de uma atividade, indicando que um agente agiu em nome de outro.
	<i>prov:InstantaneousEvent</i> (Evento instantâneo)	Representa um evento que ocorre em um único instante de tempo, sem duração, indicando um ponto específico em que algo aconteceu em uma atividade ou entidade.
	<i>prov:Role</i> (Função)	Representa a função desempenhada por uma entidade ou agente em um contexto específico, indicando o papel que algo ou alguém tem em uma atividade ou relação.
Propriedades	<i>prov:wasInfluencedBy</i> (foi Influenciado Por)	Relaciona uma entidade, atividade ou agente influenciado por outra entidade, atividade ou agente.
	<i>prov:qualifiedInfluence</i> (Influência qualificada)	Detalha uma influência específica entre entidades, atividades ou agentes, fornecendo informações adicionais.
	<i>prov:qualifiedGeneration</i> (Geração qualificada)	Detalha como uma entidade foi gerada por uma atividade, incluindo detalhes específicos.

<i>prov:qualifiedDerivation</i> (Derivação qualificada)	Detalha a derivação de uma entidade a partir de outra, incluindo as condições ou processos envolvidos.
<i>prov:qualifiedPrimarySource</i> (Fonte Primária qualificada)	Fornecer detalhes sobre a fonte primária de uma entidade, incluindo as circunstâncias de sua origem.
<i>prov:qualifiedQuotation</i> (Citação qualificada)	Especifica detalhes sobre uma citação de uma entidade a partir de outra, incluindo o contexto da citação.
<i>prov:qualifiedRevision</i> (Revisão qualificada)	Descreve a revisão de uma entidade, incluindo detalhes sobre as modificações realizadas.
<i>prov:qualifiedAttribution</i> (Atribuição qualificada)	Relaciona uma entidade a um agente que é responsável por ela, incluindo detalhes sobre a responsabilidade.
<i>prov:qualifiedInvalidation</i> (Invalidação qualificada)	Fornecer detalhes sobre o processo e as condições sob as quais uma entidade foi invalidada.
<i>prov:qualifiedStart</i> (Início qualificado)	Detalha o início de uma atividade, incluindo o agente ou condições que iniciaram a atividade.
<i>prov:qualifiedUsage</i> (Uso qualificado)	Especifica como uma entidade foi usada em uma atividade, incluindo detalhes do uso.
<i>prov:qualifiedCommunication</i> (Comunicação qualificada)	Detalha a comunicação de informações de uma entidade para outra, incluindo o meio e o contexto da comunicação.
<i>prov:qualifiedAssociation</i> (Associação qualificada)	Descreve a relação entre uma atividade e um agente associado a ela, incluindo a natureza da associação.
<i>prov:qualifiedEnd</i> (Término qualificado)	Fornecer detalhes sobre o término de uma atividade, incluindo o agente ou condições que encerraram a atividade.
<i>prov:qualifiedDelegation</i> (Delegação qualificada)	Detalha a delegação de responsabilidades de um agente para outro, incluindo o escopo e os termos da delegação.
<i>prov:influencer</i> (influenciador)	Identifica uma entidade, atividade ou agente que influenciou outra entidade, atividade ou agente.

	<i>prov:entity</i> (entidade)	Relaciona uma entidade em uma descrição de qualificação, como parte de uma influência, derivação etc.
	<i>prov:hadUsage</i> (teve Uso)	Indica o uso de uma entidade em uma atividade, detalhando a natureza desse uso.
	<i>prov:hadGeneration</i> (Geração envolvida)	Relaciona uma entidade gerada a uma atividade, detalhando como essa geração ocorreu.
	<i>prov:activity</i> (atividade)	Representa uma atividade que ocorreu, detalhando seu contexto e resultados.
	<i>prov:agent</i> (agente)	Indica um agente envolvido em uma atividade ou entidade, detalhando seu papel.
	<i>prov:hadPlan</i> (Plano envolvido)	Descreve um plano ou estratégia relacionado a uma atividade, detalhando suas metas e processos.
	<i>prov:hadActivity</i> (Atividade envolvida)	Relaciona uma atividade a outras atividades, detalhando sua relação e influências.
	<i>prov:atTime</i> (no Tempo)	Especifica um momento específico no tempo relacionado a uma atividade ou evento.
	<i>prov:hadRole</i> (Papel envolvido)	Define o papel desempenhado por uma entidade, atividade ou agente em uma relação específica.

Fonte: W3C (2013).

A análise das classes e propriedades do PROV-O destaca as complexas interações entre agentes, atividades e entidades no contexto da proveniência de dados. Dada essa complexidade, faz-se necessário utilizar uma metodologia que permita mapear e comparar essas estruturas com os elementos descritos no padrão RDA.

Na próxima seção, será apresentada a metodologia adotada neste estudo, com foco no método *Crosswalk*, que foi utilizado para mapear as correspondências entre os elementos dos padrões RDA e PROV-O, visando analisar sua interoperabilidade e aderência aos princípios de proveniência de dados.

3 METODOLOGIA

Nesta seção, estão os métodos e procedimentos adotados para a realização do estudo sobre a integração do padrão *Resource Description and Access* (RDA) com as práticas de proveniência de dados, conforme recomendadas pela ontologia PROV do *World Wide Web Consortium* (W3C). A metodologia adotada visa garantir uma análise rigorosa e detalhada dos aspectos teóricos e práticos dessa integração.

3.1 Estrutura Metodológica

3.1.1 Abordagem da pesquisa

A pesquisa adotou uma abordagem qualitativa, considerando a complexidade e a natureza exploratória do tema. Essa abordagem permite uma análise detalhada das características, relações e implicações da integração entre RDA e PROV, proporcionando uma análise que não seria facilmente obtida por métodos quantitativos.

A abordagem qualitativa permite uma exploração dos elementos descritivos e interpretativos, para compreender como esses padrões podem ser mapeados, comparados e integrados, considerando suas respectivas complexidades.

A natureza exploratória da pesquisa justifica-se pelo objetivo de investigar a aderência entre as diretrizes da RDA e os requisitos de proveniência estabelecidos pela PROV, permitindo identificar as correspondências diretas, as semelhanças e divergências que podem surgir ao tentar alinhar esses dois padrões.

3.1.2 Método *Crosswalk*

Para comparar os elementos dos padrões RDA e PROV, utilizou-se o método *Crosswalk*, o qual envolve a identificação, o mapeamento e a análise das correspondências entre os metadados dos dois padrões, permitindo avaliar a compatibilidade e a similaridade entre eles.

O método *Crosswalk* é utilizado para mapear os elementos, semântica e sintaxe de um esquema de metadados para outro. Esse método é particularmente eficaz para promover a interoperabilidade entre esquemas de metadados, permitindo que coleções diversas sejam pesquisadas simultaneamente com uma única consulta, como se fossem um único banco de dados (interoperabilidade semântica) (CHAN; ZENG, 2006).

Conforme apontado por Chan e Zeng (2006), o método *Crosswalk* é o mais comumente utilizado para permitir a interoperabilidade entre e entre esquemas de metadados, sendo essencial na conversão e troca de metadados em ambientes digitais.

Existem duas abordagens principais para a prática do *Crosswalking*, o *Crosswalk* absoluto, o qual exige um mapeamento exato entre os elementos envolvidos, que garante a equivalência dos elementos, mas que pode deixar de fora valores de dados que não têm correspondência exata, especialmente ao mapear de um esquema mais complexo para um mais simples, e o *Crosswalk* relativo, o qual mapeia todos os elementos em um esquema de origem para pelo menos um elemento em um esquema-alvo, independentemente da equivalência semântica exata, sendo mais eficaz ao mapear de um esquema complexo para um mais simples (CHAN; ZENG, 2006).

A utilização do método *Crosswalk* apresenta desafios relacionados aos diferentes graus de equivalência entre os elementos dos esquemas mapeados. Sendo eles: um para um, um para muitos, muitos para um e um para nenhum (ZENG; XIAO, 2001, apud CHAN; ZENG, 2006).

De acordo com Arakaki e Santos (2021), o método *Crosswalk* é essencial para assegurar que as informações de proveniência sejam mantidas e corretamente mapeadas entre diferentes padrões de metadados utilizados em ambientes de bibliotecas.

3.1.3 Coleta de dados

A coleta de dados foi realizada por meio de uma revisão bibliográfica, a qual compreende as normas e recomendações emitidas pelo W3C sobre a ontologia PROV. Em seguida, foi utilizada a documentação oficial da RDA, acessível por meio do RDA Toolkit (<https://access.rdatoolkit.org/>). Cabe destacar que o acesso a essa documentação não é gratuita, exige cadastro, e é limitado a um período de teste gratuito de apenas 30 dias. Em contraste, a documentação relacionada à PROV está disponível de forma aberta e gratuita através do site oficial do W3C (<https://www.w3.org/TR/prov-o/>), o que facilita e democratiza o acesso.

Além dessas fontes, foram consultados artigos acadêmicos e publicações relevantes sobre catalogação, RDA, PROV e proveniência de dados, com o objetivo de proporcionar um embasamento teórico à pesquisa. Adicionalmente, os sites oficiais tanto da RDA quanto do

PROV foram explorados para garantir o acesso às informações mais atualizadas e detalhadas sobre as respectivas normas e práticas.

3.1.4 Procedimentos de análise

Os procedimentos de análise deste estudo seguiram uma metodologia precisa, começando pela listagem dos elementos mais relevantes dos padrões RDA e PROV. Foram identificados e selecionados os elementos de metadados do padrão RDA, bem como os elementos do PROV-O *Starting Point Terms* e do PROV-O *Expanded Terms*, que serviram como base para as comparações.

Em seguida, foi realizada uma comparação detalhada entre os elementos da RDA e do PROV-O utilizando o método *Crosswalk*. Este comparativo focou em identificar semelhanças e diferenças entre os dois padrões, com especial atenção ao grau de correspondência entre eles. A análise do grau de correspondência avaliou a correspondência entre os padrões, classificando como um-para-um, muitos-para-um ou nenhum-para-um, de modo a analisar a eficácia da RDA em incorporar os princípios de proveniência de dados estabelecidos pelo PROV. Ao longo da análise, foi feita também uma abordagem para identificar se o *Crosswalk* entre os elementos dos dois padrões era absoluto, relativo ou inexistente.

No entanto, o uso do RDA Toolkit apresentou alguns desafios, especialmente na identificação dos elementos RDA correspondentes aos do PROV-O. A limitação do acesso gratuito ao RDA Toolkit, restrito a um período de apenas 30 dias, também trouxe dificuldades adicionais ao processo de mapeamento dos termos. Devido a essas dificuldades e à limitação de tempo disponível, não foi possível realizar o mapeamento dos elementos correspondentes às *Qualified Classes and Properties* do PROV-O.

3.1.5 Instrumentos norteadores

Os instrumentos norteadores utilizados nesta pesquisa foram selecionados com o objetivo de proporcionar uma base confiável para a análise da integração entre a RDA e as práticas de proveniência de dados conforme a ontologia PROV.

A documentação oficial do W3C sobre a ontologia PROV foi uma das principais referências, oferecendo diretrizes claras e detalhadas sobre os conceitos e a aplicação da proveniência de dados em ambientes digitais.

O RDA Toolkit, por sua vez, foi utilizado como uma fonte essencial para compreender os elementos e as diretrizes do padrão RDA, apesar das limitações de acesso mencionadas anteriormente.

Além dessas fontes primárias, a pesquisa foi composta também por publicações acadêmicas que abordam sobre a RDA e o PROV-O em contextos bibliográficos e digitais.

4 ANÁLISE DE DADOS

A análise comparativa entre os elementos da RDA e as classes e propriedades da Ontologia PROV tem como objetivo identificar tanto as correspondências quanto as divergências entre esses dois padrões. Esse mapeamento é essencial para entender em que medida a RDA consegue atender aos requisitos de proveniência de dados estabelecidos pelo PROV-O.

Considerando que o padrão RDA é amplamente adotado na Catalogação para a criação de metadados e que o PROV-O é fundamental para a gestão da proveniência de dados em ambientes digitais, essa análise é interessante para compreender como a RDA integra suas práticas de proveniência para otimizar a rastreabilidade e a integridade dos dados.

As tabelas comparativas a seguir destacam as correspondências, ou a ausência delas, classificadas como absolutas, relativas ou inexistentes, conforme o nível de equivalência entre os elementos comparados.

A tabela 5, mostra uma análise comparativa entre os elementos da RDA e as classes do PROV-O *Starting Point Terms*. Ao todo, foram mapeados seis (6) elementos entre a RDA e o PROV-O em três (3) comparações. Quatro (4) elementos (66,67%) apresentaram uma correspondência direta, ou seja, *Crosswalk* absoluto. Essa categoria de correspondência indica que os conceitos ou elementos mapeados possuem uma alta equivalência entre os dois padrões.

Além disso, dois (2) elementos, ou 33,33% apresentaram uma correspondência parcial ou dependente do contexto, sendo classificado como *Crosswalk* relativo. Nesse caso, a correspondência entre os elementos da RDA e as classes do PROV-O é presente, mas pode variar dependendo do uso específico ou da interpretação adotada em diferentes contextos.

Ao analisar o grau de correspondência entre os elementos, observa-se que todos os seis (6) elementos mapeados apresentam um grau de correspondência 'um-para-um' (100%). Embora esse resultado possa indicar uma forte compatibilidade entre os elementos dos dois esquemas, na análise mais detalhada dos *Crosswalks* revela que a correspondência, apesar de direta em termos de mapeamento, nem sempre é ideal ou completa.

Esses mapeamentos sugerem que, embora exista uma relação entre os elementos da RDA e do PROV-O, a correspondência nem sempre será a ideal e pode variar de acordo com o uso ou a interpretação. Isso revela os desafios na correspondência entre os dois padrões já nas classes iniciais do PROV-O.

Tabela 5 - Comparação da RDA com as classes do PROV-O *Starting Point Terms*.

RDA	PROV	Grau de correspondência	Abordagem
<i>rda:Entity</i>	<i>prov:Entity</i>	Um-para-um	Ambos incluem entidades que representam coisas identificáveis e cobrem qualquer tipo de entidade de maneira semelhante. <i>Crosswalk: Absoluto</i>
<i>rda:Expression</i>	<i>prov:Activity</i>	Um-para-um	Na RDA é descrito como uma expressão de uma obra concretizada fisicamente ou digitalmente. Esse processo de concretização pode ser visto como uma “atividade” no sentido de que envolve a produção ou materialização de uma obra em uma forma física ou digital, o que é similar ao conceito do PROV-O que envolve a geração ou modificação de entidades. <i>Crosswalk: Relativo.</i>
<i>rda:Agent</i>	<i>prov:Agent</i>	Um-para-um	Ambos os modelos incluem entidades que atuam e podem realizar ações deliberadas. <i>Crosswalk: Absoluto.</i>

Fonte: RDA Toolkit (2024); W3C (2013).

A tabela 6 também analisa e compara os elementos da RDA e do PROV-O *Starting Point Terms* com o objetivo de avaliar o grau de correspondência entre os elementos desses dois esquemas, identificando o grau de correspondência e os tipos de *Crosswalks*.

Foram analisados 19 elementos, sendo dez (10) da RDA e nove (9) do PROV-O. O processo de mapeamento evidenciou a correspondência entre esses modelos, revelando a correspondência entre os elementos dos dois padrões da RDA e do PROV-O.

Na análise do grau de correspondência entre os elementos, observa-se que a maior parte dos mapeamentos, os quais doze (12) elementos (66,67%) apresentam um grau de “um-para-um”. Isso significa que, na maioria das vezes, há uma correspondência direta entre um elemento da RDA e um termo do PROV-O.

Em contrapartida, seis (6) elementos (22,22%) do mapeamento foram classificados como “muitos-para-um”. Nesse caso, dois ou mais termos da RDA são agregados em um único termo do PROV-O. Este tipo de correspondência pode ocorrer quando o PROV-O adota uma abordagem mais geral ou mais específica para representar conceitos que, na RDA, podem ser divididos em elementos mais específicos e detalhados.

Ademais, em um (1) elemento do PROV-O (11,11%) não foi encontrado correspondência entre os elementos da RDA, sendo classificado como “nenhum-para-um”.

Os *Crosswalks* classificados como absolutos representam a maioria, correspondendo a doze (12) elementos (66,67%) dos casos. Nesses casos, há uma correspondência direta e clara entre os elementos dos dois esquemas, permitindo uma transposição sem a necessidade de grandes ajustes.

Os *Crosswalks* parciais, que abrangem seis (6) elementos (22,22%) dos elementos, indicam correspondências que, embora presentes, não são completas ou totalmente diretas. Essas situações geralmente envolvem a combinação de múltiplos elementos da RDA para corresponder a um único termo no PROV-O, refletindo diferenças na estruturação dos conceitos entre os esquemas.

Por fim, um (1) caso, ou 11,11%, foi classificado com *Crosswalk* inexistente, indicando que não há uma correspondência direta entre os termos dos dois modelos.

Tabela 6 - Comparação da RDA com as propriedades do PROV-O *Starting Point Terms*.

RDA	PROV	Grau de correspondência	Abordagem
<i>rda: generation</i>	<i>prov:was Generated By</i>	Um-para-um	Ambos os modelos reconhecem uma relação direta entre uma atividade (ou evento) e a geração de entidades. <i>Crosswalk: Absoluto.</i>
<i>rda: derivative work relationship</i>	<i>prov:was Derived From</i>	Um-para-um	Ambos os elementos descrevem uma relação de transformação ou adaptação de uma entidade ou obra a partir de outra. <i>Crosswalk: Absoluto</i>
<i>rda: statement of responsibility</i>	<i>prov:was Attributed To</i>	Um-para-um	Ambos os elementos descrevem uma relação de transformação ou adaptação de uma entidade ou obra a partir de outra. <i>Crosswalk: Absoluto.</i>

<i>rda:beginning of</i>	<i>prov:startedAt Time</i>	Um-para-um	Ambos os elementos indicam o início de um evento ou atividade, mas a RDA é mais amplo em sua aplicação. Ambos lidam com conceitos de início temporal, mas com diferentes níveis de precisão e aplicação contextual. <i>Crosswalk: Relativo.</i>
<i>rda:source consulted</i>	<i>prov:used</i>	Um-para-um	Ambos os elementos lidam com o conceito de utilização de recursos. Na RDA é mais específico ao se referir a fontes consultadas para a criação de uma obra, enquanto no PROV-O é mais geral, abrangendo qualquer entidade usada em uma atividade. <i>Crosswalk: Relativo.</i>
-	<i>prov:was InformedBy</i>	Nenhum-para-um	Não foi encontrado termo elemento equivalente na RDA. O conceito de atividades informativas e a influência entre elas, como definido no PROV-O, não é especificamente abordado nos elementos da RDA disponíveis para descrever relações de proveniência entre atividades. <i>Crosswalk: Inexistente.</i>
<i>rda:ending of</i>	<i>prov:ended AtTime</i>	Um-para-um	Ambos os elementos indicam o fim de algo no tempo. Na RDA é aplicado a eventos ou períodos, enquanto no PROV-O é específico para atividades no contexto de proveniência. Ambos capturam o conceito de término temporal, mas em diferentes contextos e escopos. <i>Crosswalk: Relativo.</i>
<i>rda:related agent of agent</i>	<i>prov:was Associated With</i>	Muitos-para-um	Os três elementos indicam uma associação genérica entre agentes e atividades ou outras entidades, mas com diferentes aplicações e escopos. <i>Crosswalk: Relativo.</i>
<i>rda:related collective agent of agent</i>			
<i>rda:related agent of agent</i>	<i>prov:acted OnBehalfOf</i>	Muitos-para-um	Ambos os elementos descrevem relações de representação ou delegação de autoridade. A RDA utiliza esses elementos para indicar qualquer forma de associação ou atuação em nome de outro agente, incluindo tanto agentes individuais quanto coletivos. No PROV-O é mais específico, focando na delegação formal de autoridade e
<i>rda:related collective</i>			

<i>agent of agent</i>			responsabilidade, enquanto os elementos da RDA são mais gerais. <i>Crosswalk</i> : Relativo.
-----------------------	--	--	---

Fonte: RDA Toolkit (2024); W3C (2013).

Na análise comparativa da tabela 7, que compara os elementos da RDA com as classes do PROV-O *Expanded Terms* foram identificados e mapeados dez (10) elementos em sete (7) comparações, que revelam diferentes níveis de alinhamento entre os dois padrões.

Desses, seis (6) elementos apresentam uma correspondência “um-para-um”, o que representa 60% do total. Esse tipo de correspondência indica uma equivalência direta entre os elementos da RDA e as classes do PROV-O.

Por outro lado, quatro (4) elementos do PROV-O (57,14%), não apresentaram uma correspondência direta e foram classificados como ‘nenhum-para-um’. Isso indica que os conceitos presentes no PROV-O não foram encontrados em um elemento correspondente na RDA.

No que se refere ao grau de aderência, os seis (6) elementos com correspondência ‘um-para-um’ foram classificados como *Crosswalks* absolutos. Esses *Crosswalks* mostram uma correspondência direta, onde um elemento da RDA corresponde diretamente a um elemento do PROV-O.

Em contraste, quatro (4) elementos não apresentaram correspondência direta na RDA e foram classificados como *Crosswalks* inexistentes.

Tabela 7 - Comparação da RDA com as classes do PROV-O *Expanded Terms*.

RDA	PROV	Grau de correspondência	Abordagem
-	<i>prov: Collection</i>	Nenhum-para-um	O conceito de coleção como uma entidade específica não foi encontrando em nenhum elemento na RDA. Embora agrupamentos de recursos possam ser descritos, a RDA não trata esse conceito como uma classe específica. <i>Crosswalk</i> : Inexistente.
-	<i>prov:Empty Collection</i>	Nenhum-para-um	Não foi encontrado elemento RDA com o conceito de uma coleção vazia, como especificado no PROV-O. Portanto, não foi

			há elemento específico que indique uma coleção sem itens ou entidades. <i>Crosswalk: Inexistente.</i>
-	<i>prov:Bundle</i>	Nenhum-para-um	O PROV-O define uma entidade que agrupa e identifica de forma única um conjunto de declarações de proveniência. Na RDA, não foi encontrado elemento específico que trate do conceito de um agrupamento de itens como uma única entidade. <i>Crosswalk: Inexistente.</i>
<i>rda:person</i>	<i>prov:Person</i>	Um-para-um	Ambas as classes representam indivíduos humanos. Na RDA é usado para descrever autores, criadores e outros indivíduos relacionados a recursos bibliográficos, enquanto no PROV-O descreve indivíduos no contexto de proveniência. <i>Crosswalk: Absoluto.</i>
-	<i>prov:Software Agent</i>	Nenhum-para-um	O conceito especificado no PROV-O, não foi encontrado na RDA, que não possui um elemento específico que represente programas de software como agentes autônomos. <i>Crosswalk: Inexistente.</i>
<i>rda:corporate body</i>	<i>prov:Organization</i>	Um-para-um	Ambos os elementos representam organizações ou grupos de pessoas que atuam de forma coletiva com um propósito comum. <i>Crosswalk: Absoluto.</i>
<i>rda:place</i>	<i>prov:Location</i>	Um-para-um	Ambos os elementos representam locais ou posições específicas associadas a entidades, eventos ou atividades. <i>Crosswalk: Absoluto.</i>

Fonte: RDA Toolkit (2024); W3C (2013).

A Tabela 8 analisa e compara os elementos da RDA com as propriedades do PROV-O *Expanded Terms*, no qual foram mapeados vinte e dois (22) elementos.

Na avaliação do grau de correspondência, observou-se que seis (6) elementos (27,27%), apresentaram uma correspondência 'um-para-um'. Nesses casos, ambos os padrões compartilham uma semântica semelhante, permitindo uma correspondência direta e clara.

Por outro lado, a maioria das propriedades do PROV-O *Expanded Terms*, dezesseis (16) no total (72,73%), não apresentaram correspondência direta, sendo classificados como

‘nenhum-para-um’. Isso mostra que muitos dos elementos do PROV-O *Expanded Terms*, não têm correspondentes específicos na RDA. Essa ausência de correspondência expõe diferenças estruturais consideráveis entre os dois padrões, indicando a necessidade de maior atenção ao alinhar esses elementos.

Além do grau de correspondência, os tipos de *Crosswalk* complementa a análise dessas correspondências. Dos elementos com correspondência ‘um-para-um’, seis (6) foram classificados como *Crosswalks* absolutos, representando 27,27% do total. Esses elementos demonstram uma correspondência direta.

Entretanto, seis (6) elementos, (27,27%), foram classificados como *Crosswalks* relativos. Nesses casos, a correspondência entre os elementos da RDA e as propriedades do PROV-O depende do contexto específico, como observado na relação entre a variabilidade no uso e na interpretação desses elementos mostra a complexidade de mapeá-los de forma coerente entre os dois padrões.

Nos dez (10) elementos restantes (45,45%), não foram encontradas correspondências na RDA e foram classificados como *Crosswalks* inexistentes, indicando que não há correspondência direta entre esses elementos da RDA e as propriedades do PROV-O *Expanded Terms*.

Tabela 8 - Comparação da RDA com as propriedades do PROV-O *Expanded Terms*.

RDA	PROV	Grau de correspondência	Abordagem
<i>rda:related work of agent</i>	<i>prov:alternateOf</i>	Um-para-um	Ambos os elementos descrevem a relação entre entidades que são alternativas ou têm alguma equivalência, especialmente em contextos em que um agente está envolvido em múltiplas versões ou edições de uma obra. Ambos lidam com conceitos de alternância ou equivalência, mas em contextos e com especificidades diferentes. <i>Crosswalk: Relativo.</i>
<i>rda:related expression of work</i>	<i>prov:specialization Of</i>	Um-para-um	Ambos os elementos descrevem uma relação de especialização ou especificidade entre entidades. Na RDA representa diferentes formas ou expressões de uma obra, enquanto no PROV-O reflete uma

			relação hierárquica entre uma entidade geral e sua especialização. <i>Crosswalk: Relativo</i>
<i>rda:date of production</i>	<i>prov: generated AtTime</i>	Um-para-um	Ambos os elementos indicam o tempo em que uma entidade foi gerada ou produzida. Na RDA refere-se à data de produção de uma manifestação, enquanto no PROV-O especifica o tempo exato de geração da entidade. <i>Crosswalk: Absoluto.</i>
<i>rda:source consulted</i>	<i>prov:had Primary Source</i>	Um-para-um	Ambos os elementos descrevem a relação entre uma obra ou expressão e a fonte primária que a influenciou ou serviu de base. Na RDA refere-se à fonte consultada para criar uma obra, enquanto no PROV-O se concentra na principal fonte de evidência ou informação. <i>Crosswalk: Relativo.</i>
-	<i>prov:value</i>	Nenhum-para-um	Não foi encontrado elemento na RDA que corresponda diretamente ao elemento do PROV-O. A RDA não trata especificamente da atribuição de valores a entidades. <i>Crosswalk: Inexistente.</i>
-	<i>prov:was QuotedFrom</i>	Nenhum-para-um	Não foi encontrado elemento na RDA que represente a relação de citação como definida no PROV-O. <i>Crosswalk: Inexistente.</i>
<i>rda:revision of</i>	<i>prov:was RevisionOf</i>	Um-para-um	Ambos os elementos descrevem uma relação de revisão entre entidades. Na RDA refere-se a uma obra ou expressão que foi revisada a partir de uma versão anterior, enquanto no PROV-O representa a mesma ideia de revisão, modificação ou atualização. <i>Crosswalk: Absoluto.</i>
-	<i>prov: invalidated AtTime</i>	Nenhum-para-um	O conceito de invalidação de uma entidade não é explicitamente tratado na RDA. Não foi encontrado um elemento específico para indicar quando uma entidade deixa de ser válida ou existente. <i>Crosswalk: Inexistente.</i>
-	<i>prov:was Invalidated By</i>	Nenhum-para-um	O conceito de invalidação de uma entidade, conforme definido no PROV-O, não foi encontrado na RDA. Não foi encontrado um elemento específico que represente a o fim

			ou a invalidação de uma entidade por uma atividade. <i>Crosswalk: Inexistente.</i>
-	<i>prov:had Member</i>	Nenhum-para-um	Não foi encontrado um elemento RDA que corresponda diretamente à propriedade do PROV-O. <i>Crosswalk: Inexistente.</i>
-	<i>prov:was StartedBy</i>	Nenhum-para-um	Não foi encontrado elemento RDA que possua diretamente o conceito de uma atividade ser iniciada pela ação de uma entidade, como descrito no PROV-O. <i>Crosswalk: Inexistente.</i>
-	<i>prov:was EndedBy</i>	Nenhum-para-um	Não foi encontrado elemento RDA que corresponda diretamente à propriedade do PROV-O. <i>Crosswalk: Inexistente.</i>
-	<i>prov: invalidated</i>	Nenhum-para-um	Não foi encontrado elemento RDA que trate do conceito de invalidação de uma entidade ou atividade como especificado no PROV-O. <i>Crosswalk: Inexistente.</i>
-	<i>prov: influenced</i>	Nenhum-para-um	Não foi encontrado elemento RDA correspondente para a propriedade PROV-O. A RDA não possui uma classe ou elemento que trate explicitamente da influência entre entidades. <i>Crosswalk: Inexistente.</i>
<i>rda:location</i>	<i>prov:at Location</i>	Um-para-um	Ambos os elementos descrevem a localização de uma entidade ou evento. Na RDA refere-se ao lugar ou posição geográfica associada a uma entidade, enquanto no PROV-O especifica a localização exata no contexto de proveniência. <i>Crosswalk: Absoluto.</i>
-	<i>prov: generated</i>	Nenhum-para-um	Não foi encontrado elemento na RDA com o conceito do PROV-O de uma entidade foi gerada como resultado de uma atividade. <i>Crosswalk: Inexistente.</i>

Fonte: RDA Toolkit (2024); W3C (2013).

A análise comparativa realizada entre os elementos da RDA e os do PROV-O buscou compreender a extensão em que a proveniência, conforme estabelecida na RDA, pode ser representada através do PROV-O.

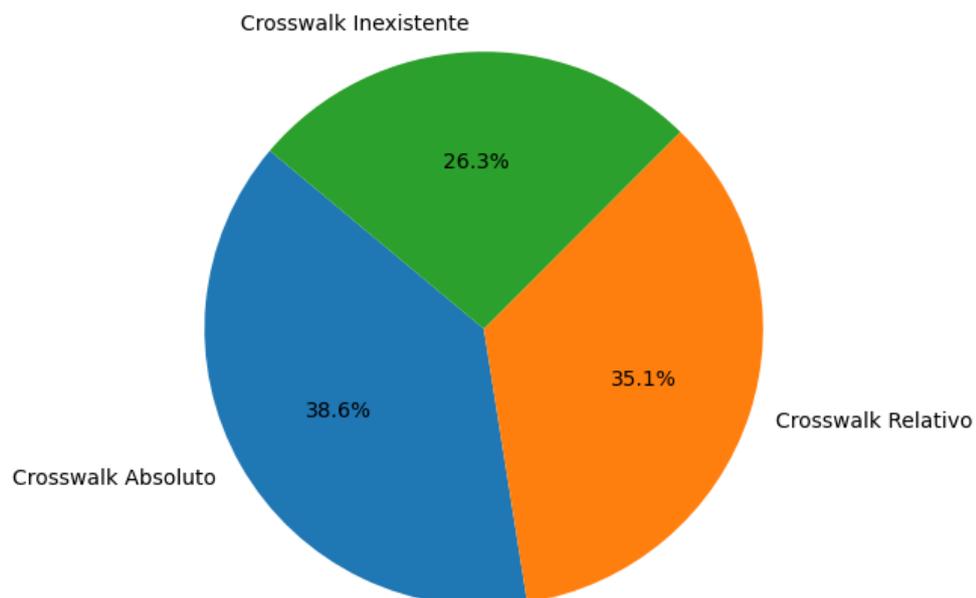
Dos cinquenta e sete (57) elementos analisados, trinta e seis (36), ou 63,15% apresentaram correspondência direta ‘um-para-um’, indicando uma boa compatibilidade entre os dois padrões para esses elementos. No entanto, seis (6) elementos (10,52%) foram classificados como ‘muitos-para-um’, o que mostra que múltiplos elementos da RDA podem ser representados em um único termo no PROV-O.

Quinze (15) elementos (26,31%) não apresentaram correspondência direta, sendo categorizados como ‘nenhum-para-um’. Essa ausência de correspondência reflete algumas lacunas na RDA em relação ao PROV-O.

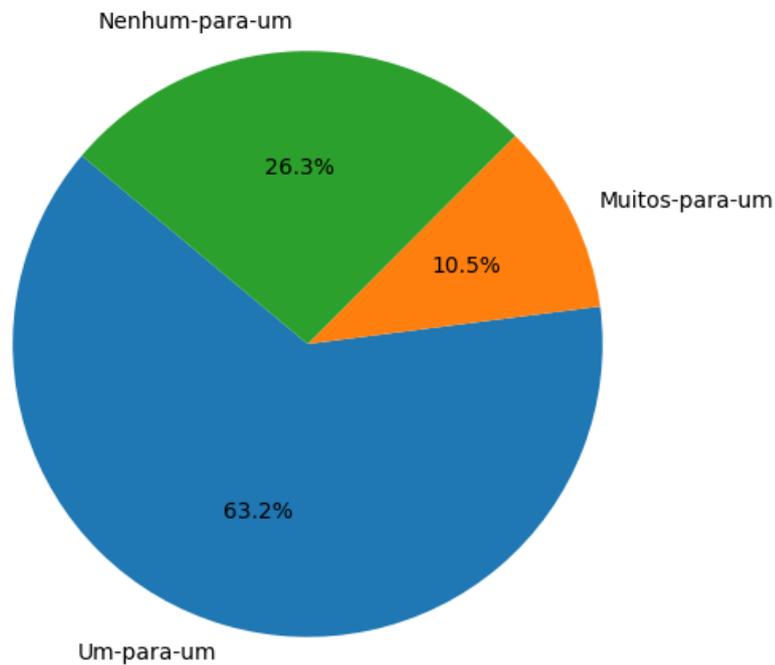
Quanto aos tipos de *Crosswalk*, vinte e dois (22) elementos (38,59%) foram classificados como *Crosswalks* absolutos, demonstrando uma correspondência direta, enquanto vinte (20) elementos (35,09%) foram identificados como *Crosswalks* relativos, dependentes do contexto específico. E por fim, quinze (15) elementos (26,32%) foram classificados como *Crosswalks* inexistentes.

Os gráficos abaixo ilustram a distribuição das correspondências e dos tipos de *Crosswalks* identificados na análise.

Gráfico 1 - Grau de correspondência entre RDA e PROV-O.



Fonte: Elaboração própria.

Gráfico 2 - Tipos de *Crosswalk* entre RDA e PROV-O.

Fonte: Elaboração própria.

Os gráficos apresentados ilustram a compatibilidade e os desafios no alinhamento entre a RDA e o modelo PROV-O. Embora exista uma correspondência significativa entre os dois padrões, áreas críticas ainda requerem maior integração para garantir uma interoperabilidade completa, especialmente no contexto da proveniência de dados em ambientes digitais.

As lacunas identificadas nos *Crosswalks* indicam que algumas das práticas de descrição da RDA precisam ser revisadas ou adaptadas para acomodar melhor os princípios da ontologia PROV. Essa análise reforça a necessidade de uma abordagem contínua e colaborativa entre bibliotecários e especialistas em dados para refinar a aplicação da RDA com vistas à proveniência de dados. Estudos futuros podem focar em preencher essas lacunas e otimizar a integração desses padrões.

5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo investigar a integração entre o padrão *Resource Description and Access* (RDA) e as práticas de proveniência de dados, conforme recomendadas pela ontologia PROV do *World Wide Web Consortium* (W3C). A pesquisa adotou uma metodologia qualitativa, no qual foi utilizado o método *Crosswalk* para mapear e comparar os elementos dos dois padrões, com o intuito de avaliar sua interoperabilidade.

O estudo trouxe a compreensão de como a RDA pode ser alinhada com as diretrizes da ontologia PROV, particularmente no que se refere à transparência, rastreabilidade e integridade das informações bibliográficas e de patrimônio cultural em contextos digitais.

Com base nos resultados, conclui-se que a RDA contempla alguns elementos de proveniência, mas de forma limitada. Embora ofereça diretrizes para rastrear a autoria, responsabilidade e algumas relações entre os recursos, sua abordagem atual não é suficiente para garantir integralmente a rastreabilidade detalhada dos dados, conforme os padrões exigidos pela ontologia PROV. Para que a RDA contemple completamente os elementos de proveniência, são necessários ajustes que permitam capturar todas as transformações e contextos pelos quais os dados passam ao longo do tempo.

O mapeamento realizado entre os elementos da RDA e o PROV-O revelou que 63,15% dos elementos mapeados apresentaram uma correspondência direta 'um-para-um' com os elementos do PROV-O, indicando um alinhamento considerável. Contudo, foi identificada uma limitação em 26,31% dos elementos, evidenciando que a RDA não abrange completamente os aspectos de proveniência modelados pela ontologia PROV.

O mapeamento permitiu verificar como os elementos de proveniência da RDA podem ser ajustados e aplicados para garantir uma representação mais precisa e eficaz dos dados, assegurando que as informações sobre a proveniência sejam preservadas e compreendidas em diferentes sistemas de informação.

Apesar das evoluções da RDA, especialmente após o Projeto 3R, sua abordagem ainda carece de alguns mecanismos para capturar as transformações e o contexto completo pelos quais os dados passam ao longo do tempo. Isso evidencia a necessidade de aprimoramentos e uma análise mais detalhada para garantir um alinhamento maior com os padrões de proveniência recomendados pela W3C.

Entre os desafios encontrados, o acesso ao RDA Toolkit, que é uma ferramenta paga, impactou a eficiência do estudo. O acesso gratuito, limitado a trinta dias, dificultou a identificação de correspondências entre os elementos de metadados dos dois padrões. O tempo disponível para a pesquisa também foi um fator limitador, impedindo uma análise das *Qualified Classes and Properties* da ontologia PROV, que poderiam fornecer mais informações sobre a interação entre os padrões.

Dada a complexidade e os desafios enfrentados, futuros trabalhos podem explorar as *Qualified Classes and Properties* do PROV-O, fornecendo uma visão mais detalhada sobre as correspondências entre a RDA e o PROV-O. Além disso, seria interessante investigar alternativas para melhorar a acessibilidade e usabilidade do RDA Toolkit, por meio do desenvolvimento de ferramentas que facilitem a identificação e correspondência de elementos de metadados relacionados à proveniência.

Os objetivos deste estudo foram cumpridos ao longo do trabalho. O objetivo principal, que era investigar a integração entre o padrão RDA e as práticas de proveniência de dados, foi tratado na seção 4, onde a análise de interoperabilidade foi realizada. O método *Crosswalk*, aplicado na seção 3, foi utilizado para mapear e comparar os elementos dos dois padrões, atendendo ao objetivo de avaliar sua compatibilidade. O primeiro objetivo, de explorar a estrutura e os princípios da RDA, foi atendido nas seções 2 e 3, enquanto o segundo objetivo, que envolvia analisar a ontologia PROV, foi abordado na seção 2.2.

O problema de pesquisa central deste estudo foi avaliar a aderência da RDA aos elementos de proveniência definidos pela ontologia PROV. Ele foi explorado principalmente nas seções 3 e 4, onde a análise do mapeamento e avaliação dos resultados demonstrou que, embora a RDA contemple parte dos elementos de proveniência, há limitações significativas em sua aderência completa ao modelo PROV-O, reforçando a necessidade de aprimoramentos.

A importância deste estudo se dá pela contribuição para a compreensão e o aprimoramento da interoperabilidade entre os padrões RDA e PROV-O, particularmente no contexto da proveniência de dados em ambientes digitais. Ao explorar as limitações e os alinhamentos entre esses padrões, este estudo reforça a importância de tal alinhamento, especialmente à medida que as demandas por maior transparência e integridade dos dados

crecem nos diversos campos do conhecimento, reforçando a relevância teórica e prática da RDA no campo da Biblioteconomia.

REFERÊNCIAS

- ALMIRÓN, F. et al. *Hallazgos en la comparación RDA-AACR2*. Buenos Aires: Biblioteca Nacional Mariano Moreno de la República Argentina - Grupo de trabajo sobre RDA, 2015. Disponível em: <https://www.bn.gov.ar/web/bibliotecarios/rda/comparacion-rda-aacr2.pdf>. Acesso em: 25 set. 2024.
- ARAKAKI, F. A.; SANTOS, P. L. V. A. C. Dados de biblioteca e proveniência: análise dos padrões de metadados. In: *ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 21., 2021, Rio de Janeiro. *Anais [...]*. Rio de Janeiro: ENANCIB, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/188>. Acesso em: 03 ago. 2024.
- ARAKAKI, F. A. Metadados e modelo PROV: perspectivas dos dados de proveniência em contextos digitais. *Informação & Informação, [S. l.]*, v. 25, n. 3, p. 187–211, 2020. DOI: 10.5433/1981-8920.2020v25n3p187. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/41869>. Acesso em: 01 dez. 2023.
- CHAN, L. M.; ZENG, M. L. Metadata Interoperability and Standardization – A Study of Methodology Part I: Achieving Interoperability at the Schema Level. *D-Lib Magazine*, v. 12, n. 6, 2006. Disponível em: <http://www.dlib.org/dlib/june06/chan/06chan.html>. Acesso em: 03 ago. 2024.
- KINCY, Chamy Pompey. *Making the move to RDA: a self-study primer for catalogers*. Rowman & Littlefield, 2014.
- MEY, E. S. A; GRAU, I. A.; BIAR, F. S. Resource Description and Access (RDA): prós e contras. *Revista Ibero-Americana de Ciência da Informação, [S. l.]*, v. 7, n. 1, p. 43–52, 2014. DOI: 10.26512/rici.v7.n1.2014.1847. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/1847>. Acesso em: 25 set. 2024.
- MODESTO, F. *Se a RDA é agnóstica, será o catalogador ateu?* InfoHome - OFAJ, março 2013. Disponível em: https://www.ofaj.com.br/colunas_conteudo.php?cod=738. Acesso em: 05 jun. 2024.
- MOREAU, L.; GROTH, P.; CHENEY, J.; LEBO, T.; MILES, S. The rationale of PROV. *Web Semantics: Science, Services and Agents on the World Wide Web, [s.l.]*, v. 35, p. 235-257, 2015. Disponível em: <https://doi.org/10.1016/j.websem.2015.04.001>. Acesso em: 03 ago. 2024.
- OLIVER, C. *Introducing RDA: A Guide to the Basics After 3R*. [S. l.]: American Library Association, 2021. 216 p. ISBN 9780838948613.
- RDA TOOLKIT. *Resource Description and Access*. Entities. Edinburgh: ALA Digital Reference, 2019. Disponível em: <https://access.rdatoolkit.org>. Acesso em: 17 jul. 2024.
- RDA TOOLKIT. *Resource Description and Access*. Guidance. Edinburgh: ALA Digital Reference, 2022. Disponível em: <https://access.rdatoolkit.org>. Acesso em: 17 jul. 2024.

RDA TOOLKIT. *Resource Description and Access*. Resources. Edinburgh: ALA Digital Reference, 2024. Disponível em: <https://access.rdatoolkit.org>. Acesso em: 17 jul. 2024.

WORLD WIDE WEB CONSORTIUM (W3C). *PROV-O: The PROV Ontology*. Expanded Terms. 2013. Disponível em: <https://www.w3.org/TR/prov-o/#description-expanded-terms>. Acesso em: 19 jun. 2024.

WORLD WIDE WEB CONSORTIUM (W3C). *PROV-O: The PROV Ontology*. Qualified Terms. 2013. Disponível em: <https://www.w3.org/TR/prov-o/#description-qualified-terms>. Acesso em: 19 jun. 2024.

WORLD WIDE WEB CONSORTIUM (W3C). *PROV-O: The PROV Ontology*. PROV-O at a glance. 2013. Disponível em: <https://www.w3.org/TR/prov-o/#prov-o-at-a-glance>. Acesso em: 19 jun. 2024.

WORLD WIDE WEB CONSORTIUM (W3C). *PROV-O: The PROV Ontology*. Starting Point Terms. 2013. Disponível em: <https://www.w3.org/TR/prov-o/#description-starting-point-terms>. Acesso em: 19 jun. 2024.

WORLD WIDE WEB CONSORTIUM (W3C). *PROV-O: The PROV Ontology*. The PROV-O Ontology Description. 2013. Disponível em: <https://www.w3.org/TR/prov-o/#description>. Acesso em: 19 jun. 2024.