

# METADADOS E MODELO PROV: PERSPECTIVAS DOS DADOS DE PROVENIÊNCIA EM CONTEXTOS DIGITAIS

## METADATA AND PROV MODEL: PERSPECTIVES OF DATA OF PROVENANCE IN DIGITAL CONTEXTS

Felipe Augusto Arakaki<sup>a</sup>

### RESUMO

**Introdução:** A compreensão dos dados é um dos requisitos para garantir a confiança das informações prestadas, além de possibilitar auditorias a partir das informações de quem as disponibilizou e quando foram publicadas. Nesse sentido, a persistência dos dados e o uso do modelo e do esquema PROV auxiliam na promoção do reuso dos dados, pois uma vez que estes são confiáveis podem ser utilizados e reutilizados. **Objetivo:** Diante deste cenário, o objetivo deste trabalho é apresentar a família de documentos PROV do *World Wide Web Consortium (W3C)* como alternativa para a representação da proveniência em contextos digitais. **Metodologia:** Caracteriza-se por uma pesquisa qualitativa e exploratória sobre proveniência em contextos digitais, por meio de uma investigação bibliográfica. O recorte da pesquisa abrange estudos publicados internacionalmente e no Brasil, nos idiomas em português, em espanhol e em inglês. **Resultados:** É apresentada uma breve discussão a respeito da proveniência em ambientes digitais. Posteriormente, as recomendações do W3C sobre a proveniência. **Conclusões:** A partir dos aspectos significativos para a estruturação da proveniência no contexto digital, esses dados são tidos como essenciais para garantir a confiabilidade das informações descritas, além de possibilitar a compreensão do contexto em que os dados foram disponibilizados e, ainda, o seu reuso.

**Descritores:** Metadados. Catalogação. Proveniência de dados. Informação e Tecnologia.

### 1 INTRODUÇÃO

O campo de estudo da representação da informação possui diversos aportes teórico-metodológicos na área de Ciência da Informação, nos quais a Catalogação apresenta-se como uma disciplina que busca representar e

---

<sup>a</sup> Doutor em Ciência da Informação pela Universidade Estadual Paulista (PPGCI/UNESP). Docente do curso de Biblioteconomia da Universidade de Brasília (UnB). E-mail: felipe.arakaki@unb.br.

organizar os aspectos físicos dos recursos informacionais, e ganhou novas formas e impactos advindos das tecnologias disponíveis.

Nesse ponto, autores como Castro e Santos (2007), Castro e Santos (2009), Alves (2010), Alves e Santos (2013), Arakaki, Simionato e Santos (2017), entre outros, já discutem sobre o impacto das tecnologias no processo de Catalogação, a qual é uma possibilidade de comunicação entre os usuários e os recursos informacionais. Ainda, pode ser um estudo, a preparação e a organização de codificações “[...] com base em itens existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir interseção entre as mensagens contidas nos itens e as mensagens internas dos usuários.” (MEY, 1995, p. 5).

Nessa relação entre catalogador e o recurso informacional, a Catalogação concentra-se em um processo que está para além da descrição do conteúdo e da definição dos metadados. Assim, situa-se em um processo de decisão multidimensional que busca tornar o objeto como único, a auxiliar nas ações de recuperação, de uso e reuso dos recursos informacionais.

Por sua vez, os metadados são fundamentais para a descrição dos recursos informacionais. Alves (2010) os define como atributos que representam uma entidade em um sistema da informação. São encontrados em diversas tipologias, mesmo sem um consenso na literatura de suas classificações (ARAKAKI, 2019; MAYERNIK, 2020). As tipologias mais encontradas são metadados descritivos e metadados administrativos.

Zeng e Qin (2016) destacam que os metadados descritivos são os mais explorados, devida à sua função e a necessidade de padronização. Enquanto os metadados administrativos, apesar de estarem presentes em diversas classificações, as subtipologias não são um consenso, pois ora algumas tipologias aparecem com subtipologia dos metadados administrativos, ora com tipologias independentes (ARAKAKI, 2019). Segundo Pomerantz (2015), esta falta de padronização decorre em razão de muitos metadados possuírem uma sobreposição de função, cujo objetivo principal é auxiliar no gerenciamento do recurso informacional.

Dentre a tipologia dos metadados administrativos, destacam-se os

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

Entretanto, o entendimento dos metadados de proveniência ainda possui lacunas e necessitam de estudos aprofundados. Ressalta-se que a representação da proveniência é fundamental para permitir a confiabilidade dos dados, pois apresenta informações sobre quem criou ou modificou um recurso, quais ações realizadas que modificaram um recurso, além de apresentar informações sobre o próprio recurso.

Diante desse cenário, o objetivo deste trabalho é apresentar a família de documentos PROV do *World Wide Web Consortium (W3C)* como alternativa para a representação da proveniência em contextos digitais. Caracteriza-se por uma pesquisa qualitativa e exploratória sobre proveniência em contextos digitais. Realizou-se a análise exploratória pelo estudo sobre a proveniência em diversos contextos, por meio de um levantamento bibliográfico. O recorte da pesquisa abrange estudos publicados internacionalmente e no Brasil, nos idiomas em português, em espanhol e em inglês.

Desta forma, este artigo aborda sobre os metadados de proveniência em ambientes digitais. Então, discutiu-se brevemente a respeito da proveniência em ambientes digitais. Posteriormente, as recomendações do W3C sobre a proveniência. Por fim, foram expostas as considerações.

## **2 PROVENIÊNCIA EM AMBIENTES DIGITAIS**

As discussões sobre proveniência em ambientes digitais surgiram em meados dos anos 2000, com algumas iniciativas a discutir como representar a proveniência em ambientes digitais.

Li e Sugimoto (2014, p. 149, tradução nossa) destacam que

A proveniência de metadados descreve e acompanha os agentes responsáveis, influenciando ações, eventos associados que causaram alterações nos metadados. O histórico de alterações de um esquema de metadados utilizados é um serviço primordial para acompanhar as alterações em suas instâncias. Portanto, a proveniência de um esquema de metadados é crucial para manter os metadados corretamente e legíveis para o sistema informacional.

Nesse contexto, como aponta Pomerantz (2015, p. 102-103, tradução nossa), os “metadados de proveniência são um mecanismo para fornecer dados sobre as entidades e seus relacionamentos com o recurso e com outras entidades.”. Segundo o autor, diversos padrões podem ser utilizados para representar a proveniência, veja:

Diversos esquemas de metadados de proveniência existem atualmente; a padronização que ocorreu em outros domínios e para outros casos de uso (*Dublin Core* para uso geral, *Art & Architecture Thesaurus* para objetos de arte, EXIF para imagens digitais etc.) [...]. Esses esquemas de proveniência compartilham muitas características: todos eles são compostos de conjuntos de elementos que identificam características do recurso ou de entidades que o influenciaram, e todos categorizam relacionamentos entre recursos e entidades. (POMERANTZ, 2015, p. 103, tradução nossa).

Como esquemas de metadados e modelos conceituais, identificaram-se na literatura o *Open Provenance Model* (OPM), o *Provenance, Authoring and Versioning* (PAV) e o *Provenance* (PROV).

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

Contudo, com o grupo de trabalho de proveniência do W3C e com pesquisadores que estudavam o OPM surgiu o PROV. De acordo com Moreau e Groth (2013) e Bivar *et al.* (2013, p. 2, tradução nossa), “o uso de proveniência, independente do modelo utilizado, fornece um fundamento essencial para avaliar

a autenticidade de dados, permitindo confiabilidade e reprodutibilidade.”. Para Bivar *et al.* (2013), com o surgimento do PROV há uma probabilidade de migração dos sistemas que utilizam OPM para PROV, pois o PROV é apoiado por uma instituição de peso, como o W3C.

Conforme Haynes (2018, p. 135), “o PROV é um padrão para metadados de proveniência, que é hospitaleiro para fornecer metadados de outros esquemas. É baseado em um modelo de agente, entidade e atividade [...]”. Ainda, apresenta um modelo geral para representar informações de proveniência. Ressalta-se que a proposta do padrão PROV não é abranger todas as especificidades de vários domínios, mas fornecer conjunto de metadados para garantir um mínimo de informações de proveniência aplicável a todos os domínios.

Derivados desses estudos, criaram-se diversas recomendações e notas para melhor representar e descrever a proveniência. Assim, sucederam-se publicações, as quais são conhecidas como documentos da família PROV. Enquanto o PAV é um padrão focado na questão garantia da proveniência, assim como identificação de pessoas, de organizações e suas funções. Isto é, o seu escopo está em quem criou, contribuiu e faz a curadoria dos dados, mas não oferece uma abordagem específica dos processos, conforme destacado por Ciccarese *et al.* (2013).

## 2.1 FAMÍLIA DE DOCUMENTOS PROV

A família PROV foi criada por um grupo de trabalho do W3C para discutir questões relativas à proveniência. Então, determinaram um modelo e outras definições as quais permitem o intercâmbio de informações de proveniência em ambientes heterogêneos (GROTH; MOREAU, 2013).

A família de documentos PROV conta com quatro recomendações: *The PROV Data Model* (PROV-DM), *The PROV Ontology* (PROV-O), *The Provenance Notation* (PROV-N) e *Constraints of the PROV Data Model* (PROV-CONSTRAINTS). Ainda, foram publicadas oito notas que auxiliam no mapeamento e nas informações sobre o modelo PROV.

O PROV-DM define um modelo de dados conceitual para a proveniência.

Arakaki (2019) explica que o PROV-DM “constitui a base para especificações da família PROV com estruturas centrais formando a essência de informações de procedência.”. Os documentos PROV-O, PROV-XML e PROV-N são baseados nesse modelo conceitual.

A recomendação PROV-O fornece um conjunto de classes, propriedades e restrições, as quais podem ser usadas para representar e trocar informações de procedência geradas em contextos e sistemas diferentes (LEBO; SAHOO; MCGUINNESS, 2013). O PROV-O é baseado na *Web Ontology Language* (OWL 2), cuja definição dada por Santarém Segundo e Coneglian (2016, p. 221) “[...] é uma linguagem de marcação semântica para a definição, a instanciação, a publicação e a partilha de ontologias na *World Wide Web*. OWL é desenvolvida como uma extensão do vocabulário RDF [...]”.

O PROV-N oferece uma notação de sintaxe compreensível, tanto por máquina quanto por seres humanos do modelo conceitual (PROV-DM), com o intuito de descrever as suas instâncias (MOREAU; MISSIER, 2013a). O PROV-N é usado para fornecer exemplos dentro do PROV-DM, assim como nas definições do PROV-CONSTRAINTS.

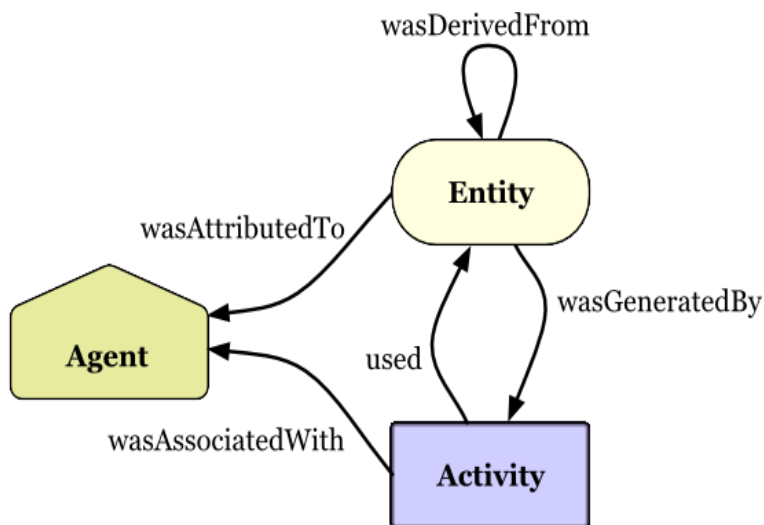
O PROV-CONSTRAINTS define um subconjunto de instâncias PROV, chamado de instâncias PROV válidas, a ajudar nas definições, nas inferências e nas restrições para outros padrões da *Web* (NIES, 2013). Ou seja, estabelece um conjunto de restrições que especifica uma noção de proveniência válida.

Projetou-se o PROV para que usuários e desenvolvedores pudessem começar com o uso básico das questões de proveniência e, posteriormente, progredir progressivamente para cenários de uso mais avançados.

## 2.2 PROV-PRIMER

O PROV-PRIMER apresenta conceitos para entender os documentos da família PRO. Assim, serve como uma introdução e uma guia para o PROV-DM e para promover o intercâmbio da proveniência na *Web* (GIL; MILES, 2013). O diagrama exposto na Figura 1 fornece uma visão geral da estrutura dos registros PROV.

**Figura 1 - Entidades PROV**



Fonte: Yolanda Gil e Miles (2013).

No PROV, *entity* (entidade) corresponde com qualquer coisa que seja físico, digital ou conceitual. Yolanda Gil e Miles (2013) esclarecem que os registros descrevem a proveniência das entidades e está pode se referir a outras.

*Activity* (atividade) é entendida como ações e processos dinâmicos. Passam a existir quando seus atributos mudam para se tornarem novas entidades. Atividades podem gerar novas entidades, por exemplo, escrever um documento o traz à existência, enquanto a revisão do documento traz uma nova versão à existência (GIL; MILES, 2013).

De acordo com Gil e Miles (2013, não paginado, tradução nossa), “um agente pode ser uma pessoa, um *software*, um objeto inanimado, uma organização ou outras entidades que podem ser responsabilizadas.”. Um agente pode assumir uma responsabilidade em uma atividade que está ocorrendo. Diante disso, considera-se que o agente se associou à atividade. Em alguns casos, um agente pode agir em nome de outro agente, isto é, um funcionário pode atuar em nome de uma organização, podendo ser expressa essa relação de responsabilidade na proveniência (GIL; MILES, 2013).

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

2013).

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

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

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

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

### **2.3 THE PROV DATA MODEL (PROV-DM)**

O PROV-DM é um modelo conceitual geral de dados, independente e extensível, o qual possibilita a interoperabilidade da proveniência entre sistemas heterogêneos. Constitui a base para especificações da família PROV com estruturas centrais, a formar a essência de informações de procedência. O modelo possui também estruturas estendidas que aprimoram e refinam as estruturas centrais para atender a usos mais avançados de proveniência. Está organizado em seis componentes: (1) entidades e atividades, e o momento em que elas foram criadas, usadas ou terminaram; (2) derivações de entidades; (3) agentes que carregam a responsabilidade das entidades que foram geradas e atividades que aconteceram; (4) um mecanismo de apoio proveniência para o empacotamento dos dados; (5) propriedades para vincular entidades que se



referem à mesma coisa; (6) coleções formando uma estrutura lógica para os seus membros (MOREAU; MISSIER, 2013a).

A estrutura base do PROV está modelada pelo diagrama *Unified Modeling Language* (UML)<sup>1</sup> e descreve o uso e a produção de entidades, atividades e agentes. Moreau e Missier (2013a, não paginado, tradução nossa) esclarecem que “uma entidade é um tipo físico, digital, conceitual ou outro tipo de coisa com alguns aspectos fixos; entidades podem ser reais ou imaginárias.”. Exemplos de entidades podem ser consideradas, porém não se limitam a um arquivo, a uma ideia ou a um objeto. “Uma atividade é algo que ocorre durante um período e age sobre ou com entidades; pode incluir o consumo, processamento, transformação, modificação, realocação, uso ou geração de entidades.” (MOREAU; MISSIER, 2013a, não paginado, tradução nossa). Exemplos de atividades são: mover, copiar, duplicar entidades digitais, entre outros. Há forte ligação entre as entidades e as atividades, a considerar que as atividades utilizam entidades e atividades produzem entidades. Enquanto o agente refere-se a

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

As relações dos modelos são: *generation* (*WasGeneratedBy*); *usage* (*Used*); *communication* (*WasInformedBy*); *derivation* (*WasDerivedFrom*); *attribution* (*WasAttributedTo*); *association* (*WasAssociatedWith*); e *delegation* (*ActedOnBehalfOf*).

O termo *generation* (geração) refere-se à conclusão do ato de produzir, ou seja, a conclusão da produção de entidade. Por outro lado, *usage* (uso) indica o início do ato de utilizar as entidades, ou seja, é o começo de usar uma entidade por uma atividade. Segundo Moreau e Missier (2013a, não paginado, tradução nossa), há uma associação entre geração e uso, pois “a geração de uma entidade por uma atividade e seu uso subsequente por outra atividade é

---

<sup>1</sup> UML é uma linguagem que auxilia na modelagem e na documentação de sistemas orientados a objetos.

denominada comunicação.” O termo *communication* (comunicação) corresponde à troca de alguma entidade por duas atividades, ou uma atividade usando alguma entidade gerada pela outra. Em alguns momentos, a utilização de uma entidade influencia a criação de outra. Logo, é considerada como uma derivação (*derivation*), ou seja, a transformação de uma entidade em outra (MOREAU; MISSIER, 2013a).

Outra relação é a atribuição (*attribution*) de responsabilidade, ou seja, a atribuição de responsabilidade a um agente. Já a associação (*association*) pode ser considerada como uma relação da atribuição de responsabilidade a um agente, indicando que o agente tinha uma função na atividade. Por fim, a delegação (*delegation*) pode ser definida como a atribuição de autoridade e responsabilidade a um agente para realizar uma atividade específica (MOREAU; MISSIER, 2013a).

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

## 2.4 PROV-O

Os editores responsáveis pelo PROV-O: *The PROV Ontology* foram Timothy Lebo, do *Rensselaer Polytechnic Institute*; Satya Sahoo, da *Case Western Reserve University*; e Deborah McGuinness, do *Rensselaer Polytechnic Institute*, nos Estados Unidos. “A Ontologia PROV (PROV-O) define a codificação da Linguagem de Ontologia da *Web* OWL2 do Modelo de Dados PROV (PROV-DM) (LEBO; SAHOO; MCGUINNESS, 2013). Dessa forma, o

PROV-O estabelece “[...] um conjunto de classes, propriedades e restrições que podem ser usadas para representar e intercambiar informações de proveniência geradas em diferentes sistemas e sob diferentes contextos.” (LEBO; SAHOO; MCGUINNESS, 2013).

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

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

A ontologia PROV possui três (3) classes principais: entidade, atividade e agente. No Quadro 1 são apresentadas as classes básicas da Ontologia PROV.

**Quadro 1 - Classes PROV-O**

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

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

Conforme exposto no Quadro 2, o PROV-O apresenta nove (9) propriedades: geração; derivação; atribuição; início; uso; comunicação; fim; associação; e delegação.

**Quadro 2 - Propriedades do PROV-O**

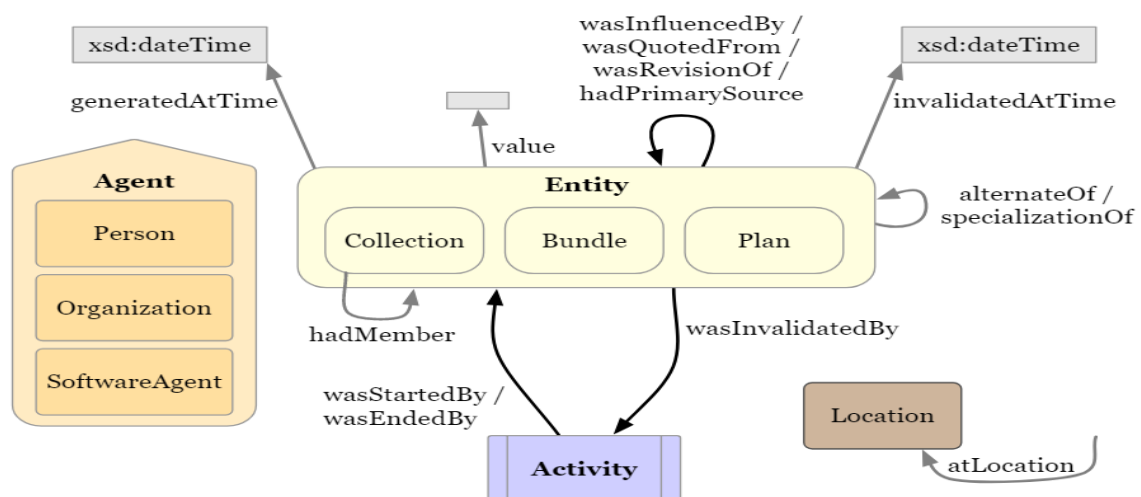
Propriedade	Definição
<i>prov:wasGeneratedBy</i> (Geração)	Foi gerado pela conclusão da produção de uma nova entidade por uma atividade.
<i>prov:wasDerivedFrom</i> (Derivação)	‘Foi derivado de’ é uma transformação de uma entidade em outra, pode ser ainda uma atualização, ou a construção de uma nova entidade baseada em uma entidade preexistente.

<i>prov:wasAttributedTo</i> (Atribuição)	'Foi atribuído para' é a atribuição de uma entidade a um agente.
<i>prov:startedAtTime</i> (Início)	O início é quando uma atividade é iniciada por uma entidade, conhecida como acionador. Qualquer uso, geração ou invalidação, envolvendo uma atividade, segue o início da atividade.
<i>prov:used</i> (Uso)	O uso é o começo de utilizar uma entidade por uma atividade.
<i>prov:wasInformedBy</i> (Comunicação)	Comunicação é a troca de uma entidade por duas atividades. Uma atividade usando a entidade gerada pela outra.
<i>prov:endedAtTime</i> (Fim)	Fim é quando uma atividade é considerada encerrada por uma entidade. Qualquer uso, geração ou invalidação envolvendo uma atividade precede o final da atividade.
<i>prov:wasAssociatedWith</i> (Associação)	Foi associado com é uma atribuição de responsabilidade a um agente de uma atividade.
<i>prov:actedOnBehalfOf</i> (Delegação)	'Agiu em nome de' ou Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica, como delegado ou representante. Enquanto o agente que atua em nome de alguma responsabilidade pelo resultado do trabalho delegado.

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

As classes e as propriedades expandidas são complementares às classes principais e muitos termos são subclasses ou subpropriedades das classes principais, conforme mostradas na Figura 2.

**Figura 2 - Diagrama das classes principais**



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

As entidades estendidas auxiliam na explicitação das classes, o que possibilita diferenças entre elas. Nesse sentido, são definidas sete (7) subclasses: coleção, coleção vazia, pacote, pessoas, agente de *software*, organização e local, conforme descritas no Quadro 3.

**Quadro 3 - Subclasses das propriedades estendidas**

Classes	Definição
<i>prov:Collection</i> (Coleção)	Uma coleção é uma entidade que fornece uma estrutura para alguns constituintes, que são entidades deles. Estes constituintes são considerados membros das coleções.
<i>prov:EmptyCollection</i> (Coleção vazia)	Uma coleção vazia é uma coleção sem membros.
<i>prov:Bundle</i> (Pacote)	Um pacote é um conjunto nomeado de descrições de proveniência, e é em si uma entidade. Permite, assim, que a proveniência da proveniência seja expressa.
<i>prov:Person</i> (Pessoa)	Indivíduo ou identidade estabelecida por um indivíduo.
<i>prov:SoftwareAgent</i> (Agente de software)	Um agente de <i>software</i> está executando o <i>software</i> .
<i>prov:Organization</i> (Organização)	Uma organização é uma instituição social ou legal, como uma empresa, sociedade etc.
<i>prov:Location</i> (Local)	Um local pode ser um local geográfico identificável (ISO 19112), mas também pode ser um local não geográfico, como um diretório, uma linha ou coluna.

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

As propriedades expandidas possuem ainda (16) subpropriedades. Elas auxiliam na representação das subclasses no PROV-O, conforme esclarecidas no Quadro 4.

**Quadro 4 - Propriedades das classes expandidas**

Propriedade	Descrição
<i>prov:alternateOf</i> (alternativa de)	Duas entidades alternativas apresentam aspectos da mesma coisa. Tais aspectos podem ser iguais ou diferentes, e as entidades alternativas podem ou não se sobrepor no tempo.
<i>prov:specializationOf</i> (especialização de)	Uma entidade que é uma especialização de outra, compartilha todos os aspectos da última e, adicionalmente, apresenta aspectos mais específicos da mesma coisa que a segunda. [...] Exemplos de aspectos incluem um período de tempo, uma abstração e um contexto associado à entidade.
<i>prov:generatedAtTime</i> (geração no tempo)	Geração é a conclusão da produção de uma nova entidade por uma atividade.
<i>prov:hadPrimarySource</i> (tem fonte primária)	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem o benefício da retrospectiva. [...] Como tal, é importante que fontes secundárias façam referência a essas fontes primárias das quais foram derivadas, para que sua confiabilidade possa ser investigada.
<i>prov:value</i> (valor)	Fornecer um valor que é uma representação direta de uma entidade.
<i>prov:wasQuotedFrom</i> (foi citado de)	'Foi citado por' é uma citação e é a repetição de (algumas ou todas) uma entidade, como texto ou imagem, por alguém que pode ou não

	ser seu autor original.
<i>prov: wasRevisionOf</i> (foi revisão de)	'Foi revisado de' é uma revisão e é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. Revisão é um caso particular de derivação.
<i>prov: invalidatedAtTime</i> (invalidado na hora)	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade.
<i>prov: wasInvalidatedBy</i> (foi invalidado por)	'Foi invalidado por' é invalidação por meio da destruição, cessação ou expiração de uma entidade existente por uma atividade.
<i>prov: hadMember</i> (tinha Membro)	Uma coleção é uma entidade que fornece uma estrutura para alguns constituintes, que são entidades deles. Estes constituintes são considerados membros das coleções.
<i>prov: wasStartedBy</i> (foi iniciado por)	'Foi iniciado por' é quando uma atividade é iniciada por uma entidade, conhecida como acionador.
<i>prov: wasEndedBy</i> (foi finalizado)	'Foi encerrado por' é quando uma atividade é encerrada por uma entidade.
<i>prov: invalidated</i> (invalidado)	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade.
<i>prov: influenced</i> (influenciado)	Influência é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
<i>prov: atLocation</i> (local)	Um local pode ser um local geográfico identificável (ISO 19112), mas também pode ser um local não geográfico, como um diretório, linha ou coluna. Como tal, há várias maneiras pelas quais a localização pode ser expressa, como por uma coordenada, endereço, marco e assim por diante.
<i>prov: generated</i> (gerado)	Geração é a conclusão da produção de uma nova entidade por uma atividade.

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

O PROV-O possui, ainda, um conjunto de vinte (20) classes e propriedades que podem ser utilizadas para fornecer atributos adicionais às categorias das classes e das propriedades básicas e estendidas. As classes definidas são: *prov:influence* (influência); *prov:entityinfluence* (influência da entidade); *prov:usage* (uso); *prov:start* (início); *prov:end* (fim); *prov:derivation* (derivação); *prov:primarysource* (fonte primária); *prov:quotation* (citação); *prov:revision* (revisão); *prov:activityinfluence* (influência da atividade); *prov:generation* (geração); *prov:communication* (comunicação); *prov:invalidation* (invalidação); *prov:agentinfluence* (influência do agente); *prov:attribution* (atribuição); *prov:association* (associação); *prov:plan* (plano); *prov:delegation*

(delegação); *prov:instantaneousevent* (evento da instanciação); e *prov:role* (papal), conforme descritas no Quadro 5.

**Quadro 5 - Classes PROV-O qualificado**

Classe	Descrição
<i>prov:Influence</i>	Influência é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
<i>prov:EntityInfluence</i>	Entidade Influenciado é a capacidade de uma entidade ter um efeito sobre o caráter, desenvolvimento ou comportamento de outra, por meio de uso, início, fim, derivação ou outro.
<i>prov:Usage</i>	O uso é o começo de utilizar uma entidade por uma atividade. Antes do uso, a atividade não havia começado a usar essa entidade e não poderia ter sido afetada pela entidade.
<i>prov:Start</i>	O início é quando uma atividade é considerada como iniciada por uma entidade, conhecida como acionador. [...] Qualquer uso, geração ou invalidação envolvendo uma atividade segue o início da atividade. Uma partida pode se referir a uma entidade que desencadeou a atividade ou a uma atividade, conhecida como acionador de partida, que gerou o acionador.
<i>prov:End</i>	Fim é quando uma atividade é considerada encerrada por uma entidade, conhecida como acionador. [...] Qualquer uso, geração ou invalidação envolvendo uma atividade precede o final da atividade. Um fim pode se referir a uma entidade desencadeadora que encerrou a atividade ou a uma atividade, conhecida como <i>ender</i> , que gerou o acionador.
<i>prov:Derivation</i>	Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova, ou a construção de uma nova entidade baseada em uma outra preexistente.
<i>prov:PrimarySource</i>	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento de seu estudo, sem o benefício da retrospectiva. É importante que fontes secundárias façam referência a essas fontes primárias das quais foram derivadas, para que sua confiabilidade possa ser investigada.
<i>prov:Quotation</i>	Uma citação é a repetição de (algumas ou todas) entidade, como texto ou imagem, por alguém que pode ou não ser seu autor original.
<i>prov:Revision</i>	Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. Revisão é um caso particular de derivação.
<i>prov:ActivityInfluence</i>	Influência da atividade é a capacidade de uma atividade de afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa por meio de geração, invalidação, comunicação ou outra.

<i>prov:Generation</i>	Geração é a conclusão da produção de uma nova entidade por uma atividade. Essa entidade não existia antes da geração e se torna disponível para uso após essa geração.
<i>prov:Communication</i>	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
<i>prov:Invalidation</i>	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade. A entidade não está mais disponível para uso (ou posterior invalidação) após a invalidação. Qualquer geração ou uso de uma entidade precede sua invalidação.
<i>prov:AgentInfluence</i>	<i>AgentInfluence</i> é a capacidade de um agente para influenciar o caráter, o desenvolvimento ou o comportamento de outro por meio de atribuição, associação, delegação ou outro.
<i>prov:Attribution</i>	Atribuição é a atribuição de uma entidade a um agente. Quando uma entidade 'e' é atribuída ao agente 'ag', a entidade 'e' foi gerada por alguma atividade não especificada que, por sua vez, foi associada ao agente 'ag'. Assim, essa relação é útil quando a atividade não é conhecida ou irrelevante.
<i>prov:Association</i>	Uma associação de atividades é uma atribuição de responsabilidade a um agente de uma atividade, indicando que o agente tinha uma função na atividade. Além disso, permite que um plano seja especificado, que é o plano pretendido pelo agente para atingir algumas metas no contexto dessa atividade.
<i>prov:Plan</i>	Um plano é uma entidade que representa um conjunto de ações ou etapas pretendidas por um ou mais agentes para atingir alguns objetivos.
<i>prov:Delegation</i>	Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que atua em nome de retém alguma responsabilidade pelo resultado do trabalho delegado.
<i>prov:InstantaneousEvent</i>	O modelo de dados PROV é implicitamente baseado em uma noção de eventos instantâneos (ou apenas eventos), que marcam transições no mundo. Os eventos incluem geração, uso ou invalidação de entidades, bem como início ou término de atividades. Essa noção de evento não é de primeira classe no modelo de dados, mas é útil para explicar seus outros conceitos e sua semântica.
<i>prov:Role</i>	Uma função é a função de uma entidade ou agente em relação a uma atividade, no contexto de um uso, geração, invalidação, associação, início e fim.

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

Foram definidas vinte e cinco (25) propriedades que podem se qualificadas no PROV-O. Elas são: *prov:wasinfluencedby* (foi influenciado por); *prov:qualifiedinfluence* (influência qualificada); *prov:qualifiedgeneration* (geração qualificada); *prov:qualifiedderivation* (derivação qualificada);



*prov:qualifiedprimarysource* (fonte primária qualificada); *prov:qualifiedquotation* (citação qualificada); *prov:qualifiedrevision* (revisão qualificada); *prov:qualifiedattribution* (atribuição qualificada); *prov:qualifiedinvalidation* (invalidação qualificada); *prov:qualifiedstart* (início qualificado); *prov:qualifiedusage* (uso qualificado); *prov:qualifiedcommunication* (comunicação qualificada); *prov:qualifiedassociation* (associação qualificada); *prov:qualifiedend* (extremidade qualificada); *prov:qualifieddelegation* (delegação qualificada); *prov:influencer* (influenciador); *prov:entity* (entidade); *prov:hadusage* (tinha uso); *prov:hadgeneration* (teve geração); *prov:activity* (atividade); *prov:agent* (agente); *prov:hadplan* (tinha plano); *prov:hadactivity* (atividade); *prov:attime* (no tempo); e *prov:hadrole* (tinha papel), conforme apresentadas no Quadro 6.

**Quadro 6 - Propriedades do PROV-O qualificado**

Propriedade	Definição
<i>prov:wasInfluencedBy</i>	'Foi influenciado por' é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa, por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
<i>prov:qualifiedInfluence</i>	Influência é a capacidade de uma entidade, atividade ou agente afetar o caráter, o desenvolvimento ou o comportamento de outra pessoa, por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
<i>prov:qualifiedGeneration</i>	Geração é a conclusão da produção de uma nova entidade por uma atividade. Essa entidade não existia antes da geração e se torna disponível para uso após essa geração.
<i>prov:qualifiedDerivation</i>	Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova, ou a construção de uma nova entidade baseada em uma entidade preexistente.
<i>prov:qualifiedPrimarySource</i>	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem o benefício da retrospectiva. Por causa da objetividade das fontes primárias, elas "falam por si" de maneiras que não podem ser capturadas através do filtro de fontes secundárias. Como tal, é importante que fontes secundárias façam referência a essas fontes primárias das quais foram derivadas, para que sua confiabilidade possa ser investigada. Uma relação de fonte primária é um caso particular de derivação de materiais secundários de suas fontes primárias. Reconhece-se

	que a determinação de fontes primárias pode ser de interpretação e deve ser feita de acordo com as convenções aceitas no domínio do aplicativo.
<i>prov:qualifiedQuotation</i>	Uma citação é a repetição de (algumas ou todas) uma entidade, como texto ou imagem, por alguém que pode ou não ser seu autor original. A citação é um caso particular de derivação.
<i>prov:qualifiedRevision</i>	Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. Revisão é um caso particular de derivação.
<i>prov:qualifiedAttribution</i>	Atribuição é a atribuição de uma entidade a um agente. Quando uma entidade 'e' é atribuída ao agente 'ag', a entidade 'e' foi gerada por alguma atividade não especificada que, por sua vez, foi associada ao agente 'ag'. Assim, essa relação é útil quando a atividade não é conhecida ou irrelevante.
<i>prov:qualifiedInvalidation</i>	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade. A entidade não está mais disponível para uso (ou posterior invalidação) após a invalidação. Qualquer geração ou uso de uma entidade precede sua invalidação.
<i>prov:qualifiedStart</i>	O início é quando uma atividade é considerada como iniciada por uma entidade, conhecida como acionador. A atividade não existia antes do seu início. Qualquer uso, geração ou invalidação envolvendo uma atividade segue o início da atividade. Uma partida pode se referir a uma entidade desencadeadora que desencadeou a atividade ou a uma atividade, conhecida como acionador de partida, que gerou o acionador.
<i>prov:qualifiedUsage</i>	O uso é o começo de utilizar uma entidade por uma atividade. Antes do uso, a atividade não havia começado a utilizar essa entidade e não poderia ter sido afetada pela entidade.
<i>prov:qualifiedCommunication</i>	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
<i>prov:qualifiedAssociation</i>	Uma associação de atividades é uma atribuição de responsabilidade a um agente de uma atividade, indicando que o agente tinha uma função na atividade. Além disso, permite que um plano seja especificado, que é o plano pretendido pelo agente para atingir algumas metas no contexto dessa atividade.
<i>prov:qualifiedEnd</i>	Fim é quando uma atividade é considerada encerrada por uma entidade, conhecida como acionador. A atividade não existe mais após o seu término. Qualquer uso, geração ou invalidação envolvendo uma atividade precede o final da atividade. Um fim pode se referir a uma entidade desencadeadora que encerrou a atividade ou a uma atividade, conhecida como <i>ender</i> , que gerou o acionador.
<i>prov:qualifiedDelegation</i>	Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante,

	enquanto o agente que atua em nome de alguém retém alguma responsabilidade pelo resultado do trabalho delegado. Por exemplo, um estudante agia em nome de seu supervisor, que agia em nome do presidente do departamento, que agia em nome da universidade; todos esses agentes são responsáveis de alguma forma pela atividade que ocorreu, mas não dizemos explicitamente quem é o responsável e em qual grau.
<i>prov:influencer</i>	Essa propriedade é usada como parte do padrão de influência qualificado. Subclasses de <i>prov: Influenciar</i> o uso dessas subpropriedades para referenciar o recurso (entidade, agente ou atividade), cuja influência está sendo qualificada.
<i>prov:entity</i>	A propriedade <i>prov: entity</i> faz referência a <i>prov: entidade</i> que influenciou um recurso. Esta propriedade aplica-se a uma <i>prov: EntityInfluence</i> , que é dada por uma <i>subproperty</i> de <i>prov: qualifiedInfluence</i> da <i>prov</i> influenciada: <i>Entity</i> , <i>prov: Activity</i> ou <i>prov: Agent</i> .
<i>prov:hadUsage</i>	O uso é o começo de utilizar uma entidade por uma atividade. Antes do uso, a atividade não havia começado a operar essa entidade e não poderia ter sido afetada pela entidade.
<i>prov:hadGeneration</i>	Geração é a conclusão da produção de uma nova entidade por uma atividade. Essa entidade não existia antes da geração e se torna disponível para uso após essa geração.
<i>prov:activity</i>	A propriedade <i>prov: activity</i> faz referência a <i>prov: Atividade</i> que influenciou um recurso. Esta propriedade aplica-se a uma <i>prov: ActivityInfluence</i> , que é dada por uma <i>subproperty</i> de <i>prov: qualifiedInfluence</i> da <i>prov</i> influenciada: <i>Entity</i> , <i>prov: Activity</i> ou <i>prov: Agent</i> .
<i>prov:agent</i>	A propriedade <i>prov: agent</i> faz referência a um <i>prov: Agente</i> que influenciou um recurso. Esta propriedade aplica-se a um <i>prov: AgentInfluence</i> , que é dada por uma <i>subproperty</i> de <i>prov: qualifiedInfluence</i> da <i>prov</i> influenciado: <i>Entity</i> , <i>prov: Activity</i> ou <i>prov: Agent</i> .
<i>prov:hadPlan</i>	Um plano é uma entidade que representa um conjunto de ações ou etapas pretendidas por um ou mais agentes para atingir alguns objetivos.
<i>prov:hadActivity</i>	Uma atividade é algo que ocorre durante um período de tempo e age sobre ou com entidades; pode incluir o consumo, o processamento, a transformação, a modificação, a realocação, o uso ou a geração de entidades.
<i>prov:atTime</i>	O modelo de dados PROV é implicitamente baseado em uma noção de eventos instantâneos (ou apenas eventos), que marcam transições no mundo. Os eventos incluem geração, uso ou invalidação de entidades, bem como início ou término de atividades. Essa noção de evento não é de primeira classe no modelo de dados, mas é útil para explicar seus outros conceitos e sua semântica.
<i>prov:hadRole</i>	Uma função é a função de uma entidade ou agente em relação a uma atividade, no contexto de um uso, geração, invalidação, associação, início e fim.

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

No Quadro 7, há exemplificação da descrição da proveniência dos dados. No exemplo, apresenta-se a descrição de duas (2) entidades, o *entity* e *agent*. A respeito da *Entity*, as propriedades *dct:issued* (linha 5), *dct:publisher* (linha 9) e *dct:creator* (linha 12) são utilizadas para dar informações sobre a origem do conjunto de dados. Na descrição de *agent* (a partir da linha 14), a propriedade *prov:actedOnBehalfOf* (linha 18) é usada para designar que “Adrian” agiu em nome da “Agência de Transporte da MyCity” (LÓSCIO; BURLE; CALEGARI, 2017).

#### Quadro 7 – Exemplo de registro com dados de proveniência

```

1 :stops-2015-05-05
2   a dcat:Dataset, prov:Entity ;
3   dct:title "Bus stops of MyCity" ;
4   dcat:keyword "transport", "mobility", "bus" ;
5   dct:issued "2015-05-05"^^xsd:date ;
6   dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
7   dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
8   dct:spatial <http://sws.geonames.org/3399415> ;
9   dct:publisher :transport-agency-mycity ;
10  dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
11  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
12  dct:creator :adrian
13
14 :adrian
15  a foaf:Person, prov:Agent ;
16  foaf:givenName "Adrian" ;
17  foaf:mbox <mailto:adrian@mycitytransport.org> ;
18  prov:actedOnBehalfOf :transport-agency-mycity
19 .
20 :transport-agency-mycity
21  a foaf:Organization, prov:Agent ;
22  foaf:name "Transport Agency of Mycity"

```

**Fonte:** Lóscio, Burle e Calegari (2017, não paginado)

A questão de proveniência também se destacou como tópicos de melhores práticas para publicação de dados na *Web* do W3C, publicado em 2017. O documento aborda catorze (14) tópicos em que, dentre eles, um que trata da importância da proveniência dos dados, tópico “8.4 *Data Provenance*”, e categorizam quais os benefícios dessas práticas para a publicação dos dados.

A proveniência é tratada na orientação BP5 “*Provide data provenance information*” e sugere-se que o fornecimento de informações completas sobre as origens dos dados e quaisquer alterações realizadas (LÓSCIO; BURLE; CALEGARI, 2017). Os benefícios dessa boa prática correspondem ao reuso, à

compreensão e à confiança.

Com as informações de proveniência descritas, possibilita-se o entendimento do contexto em que o recurso foi disponibilizado e descrito. A compreensão dos dados é um dos requisitos para garantir a confiança das informações prestadas. Ademais, possibilita auditorias, a partir de informações de quem as disponibilizou e quando foram publicadas. Por fim, a persistência dos dados e o uso do modelo e do esquema PROV auxiliam na promoção do reuso dos dados, pois uma vez que estes são confiáveis, podem ser utilizados e reusados.

### 3 CONSIDERAÇÕES

O artigo apontou a família PROV como alternativa para representar a proveniência em contextos digitais. Como visto na documentação do PROV, foi exposto as recomendações *The PROV Data Model* (PROV-DM), *The PROV Ontology* (PROV-O), *The Provenance Notation* (PROV-N) e *Constraints of the PROV Data Model* (PROV-CONSTRAINTS).

Apresentou-se um modelo geral, o qual é estruturado em três entidades básicas: recurso, atividades e agente. A proveniência do recurso está relacionada às informações básicas do próprio recurso, como foi criado e a data de criação. A proveniência no âmbito das atividades está voltada às alterações que o recurso sofreu ao longo dos anos, por exemplo, atualização e qual foi realizada, bem como quando ocorreu. Enquanto a proveniência do agente está relacionada a quem fez e é responsável pela criação, alterações e guarda do recurso. Essas são três entidades básicas para representar a proveniência. Destaca-se que, em contextos específicos, sugere-se um aprofundamento dos conceitos e metadados de proveniência para abranger domínios específicos e particulares.

Outros conceitos básicos para entender o PROV, e discutidos ao longo do texto, foram: função, derivação, revisão, atividade, plano e tempo. Cada um desses conceitos apresenta aspectos significativos para a estruturação da proveniência no contexto digital. Essas informações são essenciais para garantir a confiabilidade das informações descritas, além de possibilitar a compreensão

do contexto que os dados foram disponibilizados. Permite, ainda, o reuso dos dados.

## REFERÊNCIAS

- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação – Faculdade de Filosofia e Ciências) – Universidade Estadual Paulista “Júlio Mesquita Filho”, Marília, 2010.
- ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.
- ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados: modelo baseado na família PROV**. 2019. 139 f. Tese (Doutorado em Ciência da Informação – Faculdade de Filosofia e Ciências) – Universidade Estadual Paulista “Júlio Mesquita Filho”, Marília, 2019.
- ARAKAKI, F. A.; SIMIONATO, A. C.; SANTOS, P. L. V. A. da C. Catalogação e tecnologia: interseções com a *web* semântica. **Inf. Prof.**, Londrina, v. 6, n. 2, p. 03-19, 2017. Disponível em: <http://dx.doi.org/10.5433/2317-4390.2017v6n2p03>.
- BIVAR, B.; SANTOS, L.; KOHWALTER, T. C.; MARINHO, A.; MATTOSO, M.; BRAGANHOLO, V. *in*: BRESCI, 2013, Maceió, Brazil. **Proceedings [...]**. Maceió: [s.l.], 2013.
- CASTRO, F. F. de; SANTOS, P. L. V. A. da C. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da *web* semântica. **Informação & Sociedade: Estudos.**, João Pessoa, v. 17, n. 2, maio 2007.
- CASTRO, F. F. de; SANTOS, P. L. V. A. da C. Uso das tecnologias na representação descritiva: o padrão de descrição bibliográfica semântica *MarcOnt Initiative* nos ambientes informacionais digitais. **Ciência da Informação**, Brasília, v. 38, n. 1, p. 74-85, abr. 2009. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652009000100005&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652009000100005&lng=en&nrm=iso). Acesso em: 03 out. 2020.
- CICCARESE, P.; SOILAND-REYES, S.; BELHAJJAME, K.; GRAY, A.; GOBLE, C.; CLARK, T. PAV Ontology: Provenance, Authoring and Versioning. **Journal of Biomedical Semantics**, [s.l.], v. 4, n. 1, p. 37, 2013. Disponível em: <http://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-37>. 03 out. 2020.
- CRUZ, S. M. S. da; CAMPOS, M. L. M.; MATTOSO, M. **Towards a taxonomy of provenance in scientific workflow management systems**. [s.l.]: IEEE, 2009. p. 259–266.

GIL, Y.; MILES, S. **PROV Model Primer**. 2013. Disponível em: <https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/>. Acesso em: 03 out. 2020.

GIL, Y.; CHENEY, J.; GROTH, P.; HARTIG, O.; MILES, S.; MOREAU, L.; SILVA, P. P. da. **Provenance XG Final Report**. 2010. Disponível em: [https://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/#What\\_is\\_provenance](https://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/#What_is_provenance). Acesso em: 03 out. 2020.

GROTH, P.; MOREAU, L. **PROV-Overview**: an overview of the prov family of documents. 2013. Disponível em: <https://www.w3.org/TR/prov-overview/>. Acesso em: 10 out. 2020.

HAYNES, D. **Metadata for Information Management and Retrieval: Understanding metadata and its use**. Londres: Facet Publishing, 2018.

LEBO, T.; SAHOO, S.; MCGUINNESS, D. **PROV-O**: The PROV Ontology. 2013. Disponível em: <https://www.w3.org/TR/2013/REC-prov-o-20130430/>. Acesso em: 03 out. 2020..

LI, C.; SUGIMOTO, S. Provenance description of metadata using PROV with PREMIS for long-term use of metadata. *In*: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2014, Austin EUA. **Proceeding** [...]. EUA: DCMI, 2014. p. 147–156.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web Best Practices**. 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 03 out. 2020.

MAYERNIK, M. Metadata. *In*: HJØRLAND, B.; GNOLI, C. **Encyclopedia of Knowledge Organization**. ISKO, 2020. Disponível em: <https://www.isko.org/cyclo/metadata>. Acesso em: 30 abr. 2020.

MEY, E. S. A. **Introdução à Catalogação**. Brasília: Brique de Lemos, 1995.

MOREAU, L.; GROTH, P. Provenance: an introduction to prov. *Synthesis Lectures on the Semantic Web. Theory and Technology*, [s.l.], v. 3, n. 4, p. 1–129, 2013. Disponível em: <https://www.morganclaypool.com/doi/abs/10.2200/S00528ED1V01Y201308WBE007>. Acesso em: 03 out. 2020.

MOREAU, L.; CLIFFORD, B.; FREIRE, J.; FUTRELLE, J.; GIL, Y.; GROTH, P.; KWASNIKOWSKA, N.; MILES, S.; MISSIER, P.; MYERS, J.; PLALE, B.; SIMMHAN, J.; STEPHAN, E.; BUSSCHE, J. V. D. The Open Provenance Model Core Specification (v1.1). **Future Generation Computer Systems**, [s.l.], v. 27, n. 6, p. 743–756, jun. 2011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X10001275>. Acesso em: 03 out. 2020.

MOREAU, L.; MISSIER, P. **PROV-DM: The PROV Data Model**. 2013a. Disponível em: <https://www.w3.org/TR/2013/REC-prov-dm-20130430/>. Acesso em: 03 out. 2020.

NIES, T. de. **Constraints of the PROV Data Model**. 2013. Disponível em: <https://www.w3.org/TR/prov-constraints/>. Acesso em: 03 out. 2020.

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

SEGUNDO, J. E. S.; CONEGLIAN, C. S. *Web Semântica e Ontologias: um estudo sobre construção de axiomas e uso de inferências*. **Informação & Informação**, Londrina, v. 21, n. 2, p. 217–244, dez. 2016. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/26417/20131>. Acesso em: 24 set. 2020.

ZENG, M. L.; QIN, J. **Metadata**. Londres: Facet Publishing, 2016.

## METADATA AND PROV MODEL: PERSPECTIVES OF DATA OF PROVENANCE IN DIGITAL CONTEXTS

### ABSTRACT

**Introduction:** Understanding the data is one of the requirements to ensure the confidence of the information provided, besides enabling audits based on the information of those who made it available and when it was published. However, the persistence of the data and the use of the model and the PROV scheme benefit in promoting the reuse of the data, because once these are reliable, they can be used and reused. **Objective:** The objective of this work is to present the PROV document family of the World Wide Web Consortium (W3C) as an alternative for the representation of provenance in digital contexts. **Methodology:** It is characterized by a qualitative and exploratory research on provenance in digital contexts, through a bibliographic investigation. The research section covers studies published internationally and in Brazil, in the portuguese, spanish and english. **Results:** A brief discussion about the provenance in digital environments is presented. Subsequently, the W3C recommendations on provenance. **Conclusions:** Based on the significant aspects for the structuring of provenance in the digital context, these data are considered essential to ensure the reliability of the information described, besides enabling the understanding of the context in which the data were made available and, also, its reuse.

**Descriptors:** Metadata. Cataloging. Data Provenance. Technology and information.

## METADATOS Y EL MODELO PROV: PERSPECTIVAS DE LOS DATOS DE PROCEDENCIA EN CONTEXTOS DIGITALES

### RESUMEN

**Introducción:** Comprender los datos es uno de los requisitos para garantizar la



confianza de la información proporcionada, además de permitir auditorías basadas en la información de quienes los puso a disposición y cuándo se publicaron. En este sentido, la persistencia de los datos y el uso del modelo y el esquema PROV ayudan a promover la reutilización de los datos, ya que una vez que estos son fiables se pueden utilizar y reutilizar. **Objetivo:** En vista de este escenario, el objetivo de este trabajo es presentar la familia de documentos PROV del *World Wide Web Consortium (W3C)* como una alternativa para la representación de la procedencia en contextos digitales. **Metodología:** Se caracteriza por una investigación cualitativa y exploratoria sobre procedencia en contextos digitales, a través de una investigación bibliográfica. La sección de investigación abarca estudios publicados internacionalmente y en Brasil, en portugués, español e inglés. **Resultados:** Se presenta una breve discusión sobre la procedencia en entornos digitales. Posteriormente, las recomendaciones del W3C sobre procedencia. **Conclusiones:** Sobre la base de los aspectos significativos para la estructuración de la procedencia en el contexto digital, estos datos se consideran esenciales para garantizar la fiabilidad de la información descrita, además de permitir la comprensión del contexto en el que se ponen a disposición los datos y, también, su reutilización.

**Descriptores:** Metadatos. Catalogación. Procedencia de los datos. Información y tecnología.

**Recebido em:** 15/09/2020

**Aceito em:** 30/09/2020