W3C Working Group Note

PROV-Overview

An Overview of the PROV Family of Documents

W3C Working Group Note 30 April 2013

This version:
http://www.w3.org/TR/2013/NOTE-prov-overview-20130430/
Latest published version:
http://www.w3.org/TR/prov-overview/
Previous version:
http://www.w3.org/TR/2013/WD-prov-overview-20130312/
Editors:
Paul Groth, VU University Amsterdam
Luc Moreau, University of Southampton

Abstract

Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. The PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web. This document provides an overview of this family of documents.

Status of This Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. A list of current W3C publications and the latest revision of this technical report can be found in the W3C technical reports index at http://www.w3.org/TR/.

PROV Family of Documents

This document is part of the PROV family of documents, a set of documents defining various aspects that are necessary to achieve the vision of inter-operable interchange of provenance information in heterogeneous environments such as the Web. These

documents are listed below.

- PROV-OVERVIEW (Note), an overview of the PROV family of documents (this document);
- PROV-PRIMER (Note), a primer for the PROV data model [PROV-PRIMER];
- PROV-O (Recommendation), the PROV ontology, an OWL2 ontology allowing the mapping of the PROV data model to RDF [PROV-O];
- PROV-DM (Recommendation), the PROV data model for provenance [PROV-DM];
- PROV-N (Recommendation), a notation for provenance aimed at human consumption [PROV-N];
- PROV-CONSTRAINTS (Recommendation), a set of constraints applying to the PROV data model [PROV-CONSTRAINTS];
- PROV-XML (Note), an XML schema for the PROV data model [PROV-XML];
- PROV-AQ (Note), mechanisms for accessing and querying provenance [PROV-AQ];
- PROV-DICTIONARY (Note) introduces a specific type of collection, consisting of key-entity pairs [PROV-DICTIONARY];
- PROV-DC (Note) provides a mapping between PROV-O and Dublin Core Terms [PROV-DC];
- PROV-SEM (Note), a declarative specification in terms of first-order logic of the PROV data model [PROV-SEM];
- PROV-LINKS (Note) introduces a mechanism to link across bundles [PROV-LINKS].

Implementations Encouraged

The Provenance Working Group encourages implementation of the specifications overviewed in this document. Although work on this document by the Provenance Working Group is complete, errors may be recorded in the errata and these may be addressed in future revisions.

Please Send Comments

This document was published by the Provenance Working Group as a Working Group Note. If you wish to make comments regarding this document, please send them to public-prov-comments@w3.org (subscribe, archives). All comments are welcome.

Publication as a Working Group Note does not imply endorsement by the W3C Membership. This is a draft document and may be updated, replaced or obsoleted by other documents at any time. It is inappropriate to cite this document as other than work in progress.

This document was produced by a group operating under the 5 February 2004 W3C Patent Policy. W3C maintains a public list of any patent disclosures made in connection with the deliverables of the group; that page also includes instructions for disclosing a patent. An individual who has actual knowledge of a patent which the individual believes contains Essential Claim(s) must disclose the information in accordance with section 6 of the W3C Patent Policy.

Table of Contents

1. Introduction
2. Document Roadmap
3. Namespace
4. Additional Information
A. Change Log
    A.1 Change Log Since WD Working Draft 12 March 2013
    A.2 Acknowledgements
B. References
    B.1 Informative references

1. Introduction

This document provides a non-normative overview of the PROV Family of Documents and provides a roadmap to using them. Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. The goal of PROV is to enable the wide publication and interchange of provenance on the Web and other information systems. PROV enables one to represent and interchange provenance information using widely available formats such as RDF and XML. In addition, it provides definitions for accessing provenance information, validating it, and mapping to Dublin Core. When referring to PROV, we are referring to the entire family of documents.

The design of PROV stems from the recommendations of the Provenance Incubator Group ([PROV-XG]) which performed an extensive information gathering process including use case cataloging, requirements elicitation and a literature survey. From this process, 8 broad recommendations were defined. Summarizing, the report recommends that a provenance framework should support:

1. the core concepts of identifying an object, attributing the object to person or entity, and representing processing steps;
2. accessing provenance-related information expressed in other standards;
3. accessing provenance;
4. the provenance of provenance;
5. reproducibility;
6. versioning;
7. representing procedures;
8. and representing derivation.

PROV supports all eight of the recommendations either directly or through extensibility points.

Figure 1 shows the organization of PROV and how the documents (roughly) depend on each other. The coloring scheme corresponds to the document roadmap below.

At its core is a conceptual data model (PROV-DM), which defines a common vocabulary used to describe provenance. This is instantiated by various serializations. These serializations are used by implementations to interchange provenance. To help developers and users express valid provenance, a set of

constraints (PROV-Constraints) are defined, which can be used to implement provenance validators. This is complimented by a formal semantics (PROV-SEM). Finally, to further support the interchange of provenance, additional specifications are provided for protocols to locate and access provenance (PROV-AQ), connect bundles of provenance descriptions (PROV-Links), represent dictionary style collections (PROV-Dictionary) and define how to interoperate with the widely used Dublin Core vocabulary (PROV-DC).

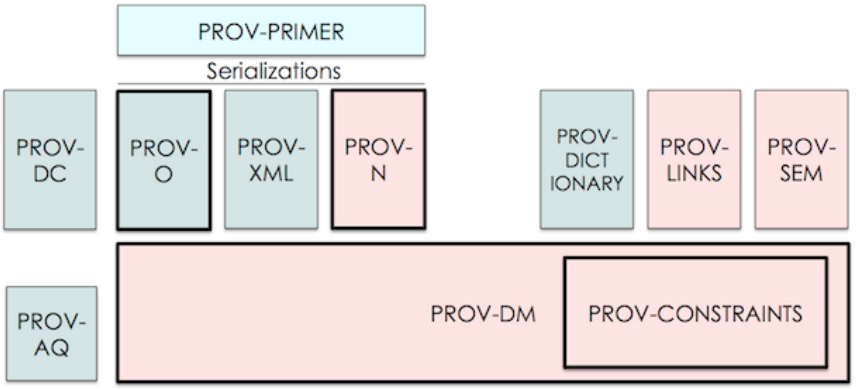

Fig. 1 The Organization of PROV

2. Document Roadmap

PROV consists of 12 documents (including this one). In order to use PROV, one need not be familiar with all of these documents. Indeed, PROV was specifically designed so that users and developers may get started quickly with basic usage and then incrementally progress to more advanced usage scenarios. To help navigate PROV, each document is broadly classified as being intended for a specific audience.

1. Users - this audience wants to understand PROV and use applications that support PROV.
2. Developers - this audience wants to develop or build applications that create and consume provenance using PROV.
3. Advanced - this audience aims to create validators, new PROV serializations, or other advanced provenance-based systems.

Figure 1 is also color coded according to this classification.

In the table below and Figure 1, we denote whether the document is a W3C Recommendation or a Working Group Note. In Figure 1, bold bordered boxes signal a W3C Recommendation.

Part	Audience	Type	Document
1	Users	Note	PROV-PRIMER is the entry point to PROV offering an introduction to the provenance data model. This is where you should start and for many may be the only document needed.
2	Developers	Rec	PROV-O defines a light-weight OWL2 ontology for the provenance data model. This is intended for the Linked Data and Semantic Web community.
3	Developers	Note	PROV-XML defines an XML schema for the provenance data model. This is intended for developers who need a native XML serialization of the PROV data model.
			PROV-DM defines a conceptual data model for provenance

4	Advanced	Rec	including UML diagrams. PROV-O, PROV-XML and PROV-N are serializations of this conceptual model.
5	Advanced	Rec	PROV-N defines a human-readable notation for the provenance model. This is used to provide examples within the conceptual model as well as used in the definition of PROV-CONSTRAINTS.
6	Advanced	Rec	PROV-CONSTRAINTS defines a set of constraints on the PROV data model that specifies a notion of valid provenance. It is specifically aimed at the implementors of validators.
7	Developers	Note	PROV-AQ defines how to use Web-based mechanisms to locate and retrieve provenance information.
8	Developers	Note	PROV-DC defines a mapping between Dublin Core and PROV-O.
9	Developers	Note	PROV-DICTIONARY defines constructs for expressing the provenance of dictionary style data structures.
10	Advanced	Note	PROV-SEM defines a declarative specification in terms of first-order logic of the PROV data model.
11	Advanced	Note	PROV-LINKS defines extensions to PROV to enable linking provenance information across bundles of provenance descriptions.

3. Namespace

All terms defined within PROV are defined within the namespace http://www.w3.org/ns/prov#. The prefix convention that is used is **prov**. Thus, no matter which document you use the namespace will be the same. The decision was made to simplify the usage of PROV.

4. Additional Information

In addition to these specifications, the PROV FAQ page addresses common questions as well as sets PROV in a broader context. This page will continue to be updated after the publication of this Note and other PROV documents. Working group members have also given several tutorials about PROV including hands-on exercises, which may be a useful place to start. In addition, one can find a variety of blog posts and web pages on PROV - a short list can be found here. For a broader review of provenance that led to the creation of PROV, there are several reports produced by the W3C Provenance Incubator group including:

1. An Overview of Provenance on the Web (slideshow - pdf)
2. Requirements for Provenance on the Web
3. State of the Art Report

Finally, the simplest way to use PROV is through one of the many applications that support it. Please see the group's implementation report [PROV-IMPLEMENTATIONS] that highlights reported software, usage in datasets, and extensions of PROV.

A. Change Log

A.1 Change Log Since WD Working Draft 12 March 2013

- Changed the status of this document section.
- Changed all URLs to PROV documents.

- Updated the figure to move prov-n outside of prov-dm and to put prov-dc on top of prov-o.
- Added section on namespaces.
- Added a paragraph discussing the broad recommendations from the incubator group.
- Editorial pass following reviews.

A.2 Acknowledgements

This document has been produced by the PROV Working Group, and its contents reflect extensive discussion within the Working Group as a whole.

Members of the PROV Working Group at the time of publication of this document were: Ilkay Altintas (Invited expert), Reza B'Far (Oracle Corporation), Khalid Belhajjame (University of Manchester), James Cheney (University of Edinburgh, School of Informatics), Sam Coppens (iMinds - Ghent University), David Corsar (University of Aberdeen, Computing Science), Stephen Cresswell (The National Archives), Tom De Nies (iMinds - Ghent University), Helena Deus (DERI Galway at the National University of Ireland, Galway, Ireland), Simon Dobson (Invited expert), Martin Doerr (Foundation for Research and Technology - Hellas(FORTH)), Kai Eckert (Invited expert), Jean-Pierre EVAIN (European Broadcasting Union, EBU-UER), James Frew (Invited expert), Irini Fundulaki (Foundation for Research and Technology - Hellas(FORTH)), Daniel Garijo (Ontology Engineering Group, Universidad Politécnica de Madrid, Spain), Yolanda Gil (Invited expert), Ryan Golden (Oracle Corporation), Paul Groth (VU University Amsterdam), Olaf Hartig (Invited expert), David Hau (National Cancer Institute, NCI), Sandro Hawke (W3C/MIT), Jörn Hees (German Research Center for Artificial Intelligence (DFKI) Gmbh), Ivan Herman, (W3C/ERCIM), Ralph Hodgson (TopQuadrant), Hook Hua (Invited expert), Trung Dong Huynh (University of Southampton), Graham Klyne (University of Oxford), Michael Lang (Revelytix, Inc.), Timothy Lebo (Rensselaer Polytechnic Institute), James McCusker (Rensselaer Polytechnic Institute), Deborah McGuinness (Rensselaer Polytechnic Institute), Simon Miles (Invited expert), Paolo Missier (School of Computing Science, Newcastle university), Luc Moreau (University of Southampton), James Myers (Rensselaer Polytechnic Institute), Vinh Nguyen (Wright State University), Edoardo Pignotti (University of Aberdeen, Computing Science), Paulo da Silva Pinheiro (Rensselaer Polytechnic Institute), Carl Reed (Open Geospatial Consortium), Adam Retter (Invited Expert), Christine Runnegar (Invited expert), Satya Sahoo (Invited expert), David Schaengold (Revelytix, Inc.), Daniel Schutzer (FSTC, Financial Services Technology Consortium), Yogesh Simmhan (Invited expert), Stian Soiland-Reyes (University of Manchester), Eric Stephan (Pacific Northwest National Laboratory), Linda Stewart (The National Archives), Ed Summers (Library of Congress), Maria Theodoridou (Foundation for Research and Technology - Hellas(FORTH)), Ted Thibodeau (OpenLink Software Inc.), Curt Tilmes (National Aeronautics and Space Administration), Craig Trim (IBM Corporation), Stephan Zednik (Rensselaer Polytechnic Institute), Jun Zhao (University of Oxford), Yuting Zhao (University of Aberdeen, Computing Science).

B. References

B.1 Informative references

[PROV-AQ]

Graham Klyne; Paul Groth; eds. *Provenance Access and Query*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-aq-20130430/

[PROV-CONSTRAINTS]

James Cheney; Paolo Missier; Luc Moreau; eds. *Constraints of the PROV Data Model*. 30 April 2013, W3C Recommendation. URL: http://www.w3.org/TR/2013/REC-prov-constraints-20130430/

[PROV-DC]

Daniel Garijo; Kai Eckert; eds. *Dublin Core to PROV Mapping*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-dc-20130430/

[PROV-DICTIONARY]

Tom De Nies; Sam Coppens; eds. *PROV Dictionary: Modeling Provenance for Dictionary Data Structures*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-dictionary-20130430/

[PROV-DM]

Luc Moreau; Paolo Missier; eds. *PROV-DM: The PROV Data Model*. 30 April 2013, W3C Recommendation. URL: http://www.w3.org/TR/2013/REC-prov-dm-20130430/

[PROV-IMPLEMENTATIONS]

Trung Dong Hynh, Paul Groth, Stephan Zednik; eds . PROV Implementation Reprot. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-implementations-20130430/'.

[PROV-LINKS]

Luc Moreau; Timothy Lebo; eds. *Linking Across Provenance Bundles*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-links-20130430/

[PROV-N]

Luc Moreau; Paolo Missier; eds. *PROV-N: The Provenance Notation*. 30 April 2013, W3C Recommendation. URL: http://www.w3.org/TR/2013/REC-prov-n-20130430/

[PROV-O]

Timothy Lebo; Satya Sahoo; Deborah McGuinness; eds. *PROV-O: The PROV Ontology*. 30 April 2013, W3C Recommendation. URL: http://www.w3.org/TR/2013/REC-prov-o-20130430/

[PROV-PRIMER]

Yolanda Gil; Simon Miles; eds. *PROV Model Primer*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/

[PROV-SEM]

James Cheney; ed. *Semantics of the PROV Data Model*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-sem-20130430.

[PROV-XG]

Yolanda Gil, James Cheney, Paul Groth, Olaf Hartig, Simon Miles, Luc Moreau, Paulo Pinherio da Silva; eds. Provenance XG Final Report. December 2010.http://www.w3.org/2005/Incubator/prov/XGR-prov-20101214/

[PROV-XML]

Hook Hua; Curt Tilmes; Stephan Zednik; eds. *PROV-XML: The PROV XML Schema*. 30 April 2013, W3C Note. URL: http://www.w3.org/TR/2013/NOTE-prov-xml-20130430/