W3C PROV: A Brief Introduction

Satya S. Sahoo
Division of Medical Informatics
Case Western Reserve University
November 13, 2014
Background: W3C and the Provenance Working Group

• World Wide Web Consortium (W3C) develops standards for the Web, such as
  - Web Design and Applications: HTML, CSS, Ajax
  - Semantic Web: RDF, OWL2, SPARQL, RIF
  - Web Services: WSDL, SOAP
  - XML Technologies: XML, XML Schema, XSLT

• Provenance Working Group (PWG) was chartered in 2012 to:
  - “…define a language for exchanging provenance information among applications.”
  - “…publication and use of provenance information of Web documents, data, and resources”
PROV Family of Specifications

• PROV is designed to model provenance information describing (from PROV Overview):
  o “Understanding how data was collected”
  o “Verifying that the process and steps used to obtain a result”
  o “Determining ownership and rights over an object”
  o “Making judgments about information to determine whether to trust it”
  o “Reproducing how something was generated”

• PROV has four primary components (W3C Recommendations):
  o PROV Data Model (PROV-DM)
  o PROV Ontology (PROV-O)
  o PROV Constraints
  o PROV Notation (PROV-N)
PROV Concepts (PROV Primer)

- **Entity**
  - wasDerivedFrom
  - used
  - wasAssociatedWith
  - wasAttributedTo
  - wasGeneratedBy

- **Agent**
  - wasAssociatedWith

- **Activity**

- **Software Agent, Person, Organization**

- **Physical, Digital or Conceptual Objects**

- **Process, Actions**
Using PROV Concepts for Data Provenance Use Case

Entity

Process

Agent

Ophthalmologist
created by
Eye Exam Report

Patient
belongs to

Eye Exam Process
generated from

Assembly of Clinical Data

Transmission of Clinical Data

Patient EHR
validated by
Primary Care Physician

S&I FRAMEWORK Use Case
Mechanism to Extend PROV Concepts for Application

• Specialization
  ○ Example: a software agent is special kind of agent

• Extended Relations
  ○ Binary relations
  ○ N-ary relations between objects: Time duration of transmission for clinical data

• Meta Provenance: Provenance of Provenance
  ○ Bundle: Set of provenance descriptions (E.g. provenance bundle of clinical document)
  ○ Collections: Provides structure to provenance descriptions
PROV Concepts (from PROV-DM)

- C1: Entities/Activities
- C2: Derivations
- C3: Agents, Responsibility, Influence
- C4: Bundles
- C5: Alternate
- C6: Collections
**PROV Relations (from PROV-DM)**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Object</th>
<th>Entity</th>
<th>Activity</th>
<th>Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>WasDerivedFrom</strong></td>
<td><strong>WasGeneratedBy</strong></td>
<td><strong>WasAttributedTo</strong></td>
</tr>
<tr>
<td>Entity</td>
<td></td>
<td>Revision</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PrimarySource</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AlternateOf</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SpecializationOf</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HadMember</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>WasInvalidatedBy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>WasAssociatedWith</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>Used</td>
<td><strong>WasStartedBy</strong></td>
<td><strong>WasInformedBy</strong></td>
<td><strong>ActedOnBehalfOf</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>WasEndedBy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Complete Family of PROV Specifications (from PROV Overview)
Conclusions

• W3C PROV specifications are designed for use in domains with large datasets for scalability and extensibility
  ○ Web, Healthcare

• Provides a good bridge to other existing W3C recommendations (e.g. XML, RDF)

• Additional Resources:
  ○ PROV Overview: http://www.w3.org/TR/prov-overview/
  ○ PROV FAQs: http://www.w3.org/2001/sw/wiki/PROV-FAQ
  ○ PROV Tools: http://www.w3.org/2001/sw/wiki/PROV
  ○ PROV Mailing List: public-prov-comments@w3.org
Acknowledgements

- W3C Provenance Working Group
- Many of the diagrams and examples used in this presentation are taken from the PROV specifications
- W3C PWG Chairs: Luc Moreau and Paul Groth
- PWG Members: Paolo Missier, Timothy Lebo, Khalid Belhajjame, Graham Klyne, Stian Soiland-Reyes, Stephan Zednic, Yolanda Gil, Simon Miles, Jun Zhao, Daniel Garijo, Helena Deus and all W3C Provenance Working Group Members