W3C Provenance Interchange Working Group Charter - Alternate Version

(Original version by Paul Groth and Luc Moreau at: http://users.ecs.soton.ac.uk/lavm/draft-charter.html,

edited by Satya Sahoo)

(November 14, 2010)

Introduction/Mission
The mission of the Provenance Interchange Working Group is to support the widespread publication and use of the provenance in Web applications, documents, data, and resources. It will publish a language (to be recommended by W3C) for exchanging provenance, and concrete specifications of the language and its inter-compatibility with existing W3C standards (XML, RDFS, OWL,...). The Provenance Interchange Working Group will leverage the work accomplished by the Provenance Incubator Group, including mappings defined between ten provenance terminologies.

1. Background
The W3C Incubator Group on Provenance has identified rapidly growing needs for provenance in social, scientific, industry, and government contexts, involving data integration across the Web and information aggregation. Provenance is unique in that it inherently draws on distributed information and thus collecting it and making sense of it require consulting different heterogeneous systems. Over time, multiple techniques to capture and represent various forms provenance have been devised, and are sometimes known under the names of lineage, pedigree, or traceability. As noted in the Incubator's state-of-the-art report, the lack of a standard model is a significant impediment to realizing such applications. It matters since provenance is key to establishing trust in documents, data, and resources. However, the Incubator's work also indicates that many provenance models exist with significantly different expressivity, fundamentally different assumptions about the system they are embedded in, but with a set of common provenance concepts.

An practical approach is to consider a core provenance language and extension mechanisms that allow any provenance model to be translated into such a lingua franca and exchanged between systems. Heterogeneous systems can then export their provenance into such a core language, and applications that need to make sense of provenance in heterogeneous systems can then import it and reason over it.
In a quest to understand emerging provenance models better, the W3C Incubator Group on Provenance decided to map from their concepts to a single target model. To this end, the Incubator Group chose to adopt the Open Provenance Model [OPM V1.1] as the target model since it is already a community model, which has undergone several revisions, and which is already adopted by 10 different systems. The Incubator group found that the emerging models for provenance, despite being originated from a wide range of domains, share a set of three primary concepts of “Process”, “Agent”, and “Artifact” as modeled in OPM. The mapping exercise by the incubator group identified that in addition to OPM core concepts, a set of additional terms that represent important provenance concepts from other provenance terminologies. These additional terms from other provenance terminologies, such as the Provenir ontology, the Proof Markup Language, and Dublin Core Metadata Initiative (DCMI), together with core OPM terms can form the bootstrapping basis of a common provenance exchange language. This community-based approach to create a will have significant advantages by drawing from the multiple existing and in-use provenance terminologies. The notable feature of OPM is that it was specifically designed as a language to exchange provenance information, and was refined and actively tested in the Provenance Challenge series, a community interoperability exercise for scientific workflows.

2. Scope

The goal of this working group is to create a common representation of provenance that can be used to exchange the provenance of documents in an interoperable manner. This representation will refine OPM along with additional terms identified by the provenance incubator group mapping work, and will be complemented with mechanisms to publish, retrieve and access provenance. The main items of work to be completed by the working group are:

- Define a language for exchanging provenance (Provenance Exchange Language) that leverages the existing provenance terminologies and the mapping work by the provenance incubator group (using OPM terms as reference).
- Define an well-defined mapping mechanism for the provenance exchange language to existing provenance terminologies in order to ensure inter-operability using standard vocabulary, for example using the OPM profiles (for Dublin Core) and use of RDFS for extending the Provenir ontology.
- Define the serialization of that exchange language in XML and map it to RDFS and OWL syntax as well as define the semantics of the overall language (as defined in RIF).

In addition to the core objectives in the scope of the Provenance Working Group, we also propose some secondary-objectives:

- Specify how to embed provenance in document with RDFa, or to access provenance held separate in services.
- Demonstrate the use of the Provenance Exchange Language in two or three exemplar use cases that entails collaboration with interested W3C groups (e.g. Health Care and Life Sciences, Social?, Sensor Web).

The specified provenance exchange language will be a refinement of a set of existing provenance terminologies starting with OPM. The refinement will adopt feedback from
the Incubator's mapping exercise, adapt terminology to avoid unnecessary technical jargon, and better characterize the notion of agent to promote inter-operability. Usage guidance will leverage a proposal on mapping Dublin Core concepts to OPM [OPM DC].

2.1 Success Criteria
- The resulting model being generated by multiple implementations including but not limited to toolkits, content management systems, workflow systems, and wikis.
- Deliver all the identified reports.

2.2 Out of Scope
- Design of a novel query language for provenance

3 Deliverables and Schedule
The Working Group has an aggressive timetable based on the premise that it builds on existing work and does not develop a new model.

3.1 Deliverables
The following deliverables have been identified. Their titles are indicative only, and the kind of targeted W3C specification has been identified in italic.
- **D1.1: Provenance Exchange Language/OPM Syntax and Abstract Model (W3C Recommendation)**. This document includes the syntax and the conceptual model.
- **D1.2: Provenance Exchange Language/OPM Semantics (W3C Note)**. Publication of this note is optional. It consists of a formal semantics of OPM.
- **D2.1: Compatibility of Provenance Exchange Language/OPM compatibility with RDF OWL (W3C Recommendation)**. Two documents define the mapping of Provenance Exchange Language/OPM syntax with RDF and OWL and also discuss the semantics of the overall language.
- **D2.2: Provenance Exchange Language/OPM XML Serialization (W3C Recommendation)**. This document specifies an XML serialization for Provenance Exchange Language/OPM that mirrors its data model.
- **D3: Provenance Exchange Language/OPM Primer (W3C Recommendation)**. Provides an easy to understand and detailed but non-normative description of the Provenance Exchange Language/OPM.
- **D5: Best Practice Cookbook (W3C Note)**. This document includes a limited set of best practice profiles that link with other relevant models, such as Dublin Core provenance-related concepts, licensing in Creative Commons, and the OpenId identity mechanism for people.
- **D6: Interoperability Guidelines (W3C Note)**. This document explains how extant provenance models can be encoded into the Provenance Exchange Language/OPM to ensure interoperable exchange of provenance across heterogeneous systems using standard vocabularies (e.g. RDFS).

3.2 Milestones
Reports will undergo the W3C development process: Working Draft (WD), Working Draft in Last Call (LC), Candidate Recommendation (CR), Proposed Recommendation (PR) and Recommendation (Rec).

Note: Need to check time constraints set by W3C development process.

<table>
<thead>
<tr>
<th>Specification</th>
<th>FPWD</th>
<th>LC</th>
<th>CR</th>
<th>PR</th>
<th>Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1.1</td>
<td>T+3</td>
<td>T+4</td>
<td>T+6</td>
<td>T+9</td>
<td>T+12</td>
</tr>
<tr>
<td>D1.2 (Optional)</td>
<td>T+9</td>
<td>T+12</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>D2.1</td>
<td>T+6</td>
<td>T+7</td>
<td>T+9</td>
<td>T+10</td>
<td>T+12</td>
</tr>
<tr>
<td>D2.2</td>
<td>T+6</td>
<td>T+7</td>
<td>T+9</td>
<td>T+10</td>
<td>T+12</td>
</tr>
<tr>
<td>D3</td>
<td>T+9</td>
<td>T+12</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>D4</td>
<td>T+6</td>
<td>T+8</td>
<td>T+10</td>
<td>T+11</td>
<td>T+12</td>
</tr>
<tr>
<td>D5</td>
<td>T+9</td>
<td>T+12</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>D6</td>
<td>T+9</td>
<td>T+12</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

3.3 Timeline View Summary

To be completed, once milestones table is finalized.

- Month T: First Teleconference
- Month T+3: D1.1 (FPWD)
- Month T+12: D1.2 (FPWD)
- Month T+6: D2.1 (FPWD)
- Month T+6: D2.2 (FPWD)
- Month T+6: D4 (WD)
- Month T+12: D3 (WD)
- Month T+12: D5 (WD)

4. Dependencies and Liaisons

- Dublin Core
- RDF working group
- RIF
- Semantic Web Health Care and Life Sciences (HCLS) Interest Group
- eGovernment working group

5. Participation

6. Communication

- Mailing List

7. Decision Policy
8. References


- [Agent1] Paul Groth and Luc Moreau, proposal for agents in OPM.
