**W3C Provenance Interchange Working Group Draft Charter**

**(October 18, 2010)**

**(First revision – November 15, 2010)**

**Second – November 15, 2010**

**Introduction/Mission**

The mission of the Provenance Interchange Working Group is to support the widespread publication and use of the provenance of Web documents, data, and resources. It will publish a W3C recommendation for a language for exchanging provenance based on W3C RDF and OWL. The language will have a graphical notation for human consumption and corresponding serialized notation for machine consumption. The Provenance Interchange Working Group will develop the new language by leveraging core concepts from existing provenance notations identified and discussed by the W3C Provenance Incubator Group..

**1. Background**

The [W3C Incubator Group on Provenance](http://www.w3.org/2005/Incubator/prov/) has identified rapidly growing needs for provenance in social, scientific, industry, and government contexts, involving data collection, integration and derivation 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, proofs, or information transformation traces. As noted in the Incubator's state-of-the-art report, the lack of a model that a large community agrees to be a standard for sharing provenance is a significant impediment to realizing the support that provenance can provide for such applications and their results. It matters since provenance is key to establishing a degree of belief in documents, data, and resources from degrees of trust related to the sources and potential transofrmations of such content. The idea that a single provenance notation representing and collecting all known dimensions and aspects of provenance that could be adopted as the internal representation language by most systems does not seem to be realistic today.

An alternative approach is to consider a provenance language that represents the core provenance concepts identified in the collection of existing provenance languages. This core should allow for extension mechanisms so that provenance notions not covered by the core may be developed and incorporated by applications that require additional expressiveness. Extension meachanisms may also be used to integrate existing notations and other standards such as Dublin Core that are related to provenance but that may not be included as part of this core provenance set. . Heterogeneous systems can then export their provenance that is expressible in this language into such a core language, and applications that need to make sense of provenance in hetereogeous systems can then import it and reason over it. Systems that require additional representational primitives may provide their own module extensions to capture notions not covered by the core.

The Incubator group found that emerging models for provenance, despite having originated from a wide range of domains and application settings, have a common set of concepts and that the establishment of such a set of core concepts is one of the first steps for the development of this language that we are provisionally calling XG. Moreover, the W3C Provenance Incubator Group has already identified the difficulty of using certain terms to name some of the core concepts and that the new language would be based on terms agreed upon the entire working group.

**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 identify core concepts and use these concepts as the foundation of the new language XG, and will complement the use of these concepts with mechanisms to publish, retrieve and access provenance. The main items of work to be completed by the working group are:

* Define a new language based on RDF/OWL for exchanging provenance that leverages the existing work on conceptual provenance models and extension mechanisms
* Specify how to embed provenance in document with RDF, or to access provenance held separate in services
* Create a primer on how to use the XG model to express the provenance of document, data, and resources
* Define extensibility mechanism of the XG model by means of profiles, and promote the wide adoption of the model through usage guidance as well as links with other models such as Dublin Core.

The specified provenance model will leverage existing conceptual models and their extension mechanism [OPM V1.1, PML2]. The naming of the concepts in the new provenance modelwill adopt feedback from the Incubator's mapping exercise, adapt terminology to avoid unnecessary technical jargon, better characterize the notion of agent to promote inter-operability by folding in proposals such as [[Agent1]](http://users.ecs.soton.ac.uk/lavm/draft-charter.html#ref-agent1) and [[Agent2]](http://users.ecs.soton.ac.uk/lavm/draft-charter.html#ref-agent2), align the notion of time with the one in the time ontology, and fine-tune the concept of profile to facilitate the extension of XG in the activities of this Working Group and beyond.

**2.1 Success Criteria**

* The resulting model being *generated* by multiple implementations including but not limited to toolkits, content management systems, workflow systems, theorem provers, mash-up systems, and wikis.
* Deliver all the identified reports.

**2.2 Out of Scope**

* Design of a novel query language for provenance
* Database provenance
* Specialization of provenance to specific application domains

**3 Deliverables and Schedule**

The Working Group has an aggressive timetable based on the premise that it builds on existing work once we have a clear understanding of the boundaries of the 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: XG Abstract Model** and RDF/OWL Serialization *(W3C Recommendation)*. This document includes includes the data model and permitted inferences using OWL and Semantic Web technologies.
* **D1.2: XG Semantics** *(W3C Note)*. Publication of this note is optional. It consists of a formal semantics of XG.
* **D3: XG Primer** *(W3C Note)* Educational/outreach material aimed at users of provenance.
* **D4: Accessing, querying provenance.** *(W3C Recommendation)*. This document specifies (1) how to access provenance from a service by means of HTTP, (3) how to query provenance through a SPARQL endpoint
* **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 XG to ensure interoperable exchange of provenance across heterogeneous systems.

**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. | | | | | |
| **Specification** | **FPWD** | **LC** | **CR** | **PR** | **Rec** |
| **D1.1** | T+3 | T+4 | T+6 | T+9 | T+12 |
| **D1.2 (Optional)** | T+9 | T+12 | n/a | n/a | n/a |
| **D2.1** | T+6 | T+7 | T+9 | T+10 | T+12 |
| **D2.2** | T+6 | T+7 | T+9 | T+10 | T+12 |
| **D3** | T+9 | T+12 | n/a | n/a | n/a |
| **D4** | T+6 | T+8 | T+10 | T+11 | T+12 |
| **D5** | T+9 | T+12 | n/a | n/a | n/a |
| **D6** | T+9 | T+12 | n/a | n/a | n/a |

**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**

* [OPM V1.1] Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric Stephan, and Jan Van den Bussche. The open provenance model core specification (v1.1). Future Generation Computer Systems, July 2010. ([doi: 10.1016/j.future.2010.07.005](http://dx.doi.org/10.1016/j.future.2010.07.005)), ([www: http://eprints.ecs.soton.ac.uk/21449/](http://eprints.ecs.soton.ac.uk/21449/)).
* [PML] Paulo Pinheiro da Silva, Deborah L. McGuinness and Richard Fikes. A Proof Markup Language for Semantic Web Services. Information Systems 31(4-5): 381-395, 2006.
* [PML2] Deborah L. McGuinness, Li Ding, Paulo Pinheiro da Silva, Cynthia Chang. PML2: A Modular Explanation Interlingua. In Proceedings of the AAAI 2007 Workshop on Explanation-aware Computing, Vancouver, British Columbia, Canada, July 22-23, 2007.
* [OPMO] The Open Provenance Model Ontology file <http://openprovenance.org/model/opmo>.
* [OPMV] Jun Zhao, the Open Provenance Model Vocabulary Specification <http://purl.org/net/opmv/ns>.
* [OPMX] Paul Groth and Luc Moreau, the Open Provenance Model XML Schema [http://openprovenance.org/model/opmo-20101012.xsd](http://openprovenance.org/model/opmx-20101012.xsd).
* [Prov Mapping] Satya Sahoo, Paul Groth, Olaf Hartig, Simon Miles, Sam Coppens, James Myers, Yolanda Gil, Luc Moreau, Jun Zhao, Michael Panzer, Daniel Garijo, Provenance Vocabulary Mappings, <http://www.w3.org/2005/Incubator/prov/wiki/Provenance_Vocabulary_Mappings>.
* [Agent1] Paul Groth and Luc Moreau, proposal for agents in OPM.
* [Agent2] Jim Myers, I Think Therefore I Am Someone Else: Understanding the confusion of granularity with Continuant/Occurrent and related perspective shifts. <http://tw.rpi.edu/portal/File:IPAW2010_ITTIA_Myers.pdf>

[OPM DC] Simon Miles, Mappi