Working Group on Provenance Solutions

Research Data Provenance IG

Last updated Tue, Dec 20, 2016 6:02:00 PM

Note: this is a generated file. Make changes to the markdown source files.

# Charter

WG Charter: A concise articulation of what issues the WG will address within a 12-18 month time frame and what its “deliverables” or outcomes will be.

## Overview

Tracking provenance for research data is vital to science and scholarship, providing answers to common questions researchers and institutions pose when sharing and exchanging data[[1]](#footnote-23):.

The tasks for this Working Group focus on finding, detailing and recommending best practices for provenance representation and management.

This group will conduct its work in the manner of a business analysis task: identifying business needs and determining solutions to business problems. Since RDA WGs are not themselves research groups (rather groups of researchers and research agencies), this group will look for existing practice and re-present that for use rather than generate new practice.

The six activity areas[[2]](#footnote-24): of the Working Group will be:

1. Common provenance Use Cases
2. Provenance patterns
3. Sharing provenance
4. Strategies for enterprise provenance management
5. Tools for provenance
6. Provenance data collections

## Deliverables

The deliverables for this Working Group are separated into four time-based cohorts as below. Short-term goals are mostly about seeking existing practice. Medium-term about determining possible output forms for the activity areas. Long-term about delivering those outputs and After-term about ensuring continuation of output custodianship, where required.

### Short-term (M6)

1. Implement a provenance Use Case recording system
2. Seek Provenance Use Cases from other IGs/WGs
3. Seek existing patterns and posit initial generalised provenance patterns
4. Seek existing provenance sharing implementations
5. Seek existing enterprise provenance management implementations
6. Seek listings of provenance tools from RDA members and the provenance research community
7. Seek existing collections of open and non-open provenance data

### Medium-term (M12)

1. De-duplicate and categorise provenance Use Cases
2. Determine possible Use Case list outputs
3. Review patterns and determine recommendations or other outputs
4. Review sharing implementations and determine recommendations or other outputs
5. Review enterprise management implementations and seek input on potential WG outputs (recommendations,recipes, etc.)
6. Review provenance tool lists and seek potential WG outputs (a consolidated list, tol descriptors, community list custodianship)
7. Review existing provenance data collections and determine how to present them to the community

### Long-term (M18)

1. Deliver Use Case list outputs
2. Deliver pattern outputs
3. Deliver sharing implementation outputs
4. Deliver enterprise management outputs
5. Deliver tool list outputs
6. Seek provenance tool list custodians
7. Deliver data collection list outputs

### After-term (M18+)

1. Hand over provenance tool and data collection listings governance to non-WG custodians

# Value Proposition

Value Proposition: A specific description of who will benefit from the adoption or implementation of the WG outcomes and what tangible impacts should result.

Effective provenance management is sought by many members of the RDA and wider science data community. We propose a working group to help those members adopt existing provenance management practice. This help will be in the form of documenting provenance use cases: centralising a list of them and generalising them to reveal common ones; documenting existing technical and business processes for provenance management, assisting organisations with sharing provenance and listing existing sources of real provenance information.

## From our P8 Slides

We plan to propose a working group on provenance solution patterns.

* The patterns should relate to core RDA interests, perhaps data/data and data/people relationships.
* Provenance vocabularies offer a level of generality/specificity that address what we perceive to be implementation gaps.
* Our goal: constructive engagement with and response to published RDA recommendations.

### WG activity points

1. **Provenance Use Cases**
   * Use Cases for provenance data or systems are often articulated in terms understood by a particular community however in our group's experience, many provenance Use Cases are differently worded instances of general Cases.
   * The establishment of a published set of UCs would allow people to compare their UCs with known UCs for which recommended implementations and other patterns may already be known.
2. **Provenance patterns**
   * Some ways of doing things in provenance are better than others. This activity is to generate provenance design patterns (for any provenance task such as representation, transmission, use etc.) perhaps in response to a series of provenance use cases that we would generate.
   * The patterns should relate to core RDA interests, perhaps data/data and data/people relationships.
3. **Sharing provenance**
   * This may only be a single class of provenance use cases but it is one that is less maturely answered by the provenance research community than, say, provenance representation. This activity might be to generate requirements for the research community to answer or perhaps find that no more research is needed for sensible recommendations for provenance sharing.
4. **Strategies for implementation**
   * Some provenance use cases apply to whole organisations (or consortia) and some organisations (or consortia) may already have experience in implementing solutions to them. This activity will list such use cases and seek descriptions of implemented or proposed solutions from members.
5. **Tools for provenance**
   * In addition to several well-known provenance conceptual models, there are tools to assist with the management of provenance. We will list those tools with comparisons in relation to RDA interests (perhaps taken from IG and other WG members).

# Engagement

Engagement with existing work in the area: A brief review of related work and plan for engagement with any other activities in the area.

In addition to serving the RDA community directly, this Working Group aims to serve the immediate interests of existing RDA groups. The reason for this is that provenance is foundational to many other RDA groups' activity and thus maximal impact on the RDA community can be achieved by aligning and assisting work in existing groups. This means this WG will engage heavily with other groups and source its primary requirements and exemplars from other groups.

## From our P8 Slides

* **Publishing Data Workflows WG**
  + **About:** "Workflows that enable persistence, quality control and access are all crucial to enhance the possibilities for greater discoverability as well as efficient and reliable reuse of research data." **and** "The objectives of this Working Group are to provide an analysis of a representative range of existing and emerging workflows and standards for data publishing, including deposit and citation, and provide reference models and implementations for application in new workflows."
  + **Intersections:**
    - Workflows that enable persistence may need a representation of the work done.
    - Discovery of workflow products.
    - Data reuse (as per Prov for Reuse Fitness Assessment).
    - Standards for data publishing ... and citation.

### Other Groups

* **Dynamic Data Citation WG**
  + Describing the conceptual model containing object relevant to dynamic data citation.
* **PID Information Types**
  + **About**: Harmonizing the basic information types associated with persistent identifiers.
  + **Intersections**:
    - Flesh out Use Case "A.10 Provenance tracing."
* **Reproducibility IG**
  + **About**: Advance and enable reproducibility in research based on or producing datasets.
  + **Intersections**: Their conception of provenance.
* **PID IG**
  + **About**: PIDs
  + **Intersection**: Requirements for PIDs to maintain provenance content (help from them).

### Other Groups

* **Archives and Records Professionals for Research Data IG**
  + **About**: Explore the areas where the principles and practices in the information disciplines of archives, records management, and research data curation overlap and where they diverge (proposed IG).
  + **Intersections**: Need for semantic understanding of archived material.
* **Data Discovery IG**
  + **About**: Findable, Accessible, Interpretable and Reusable?
  + **Intersections**: Conceptual model (upper ontology) relating elements relevant to data discovery.
* **Preservation e-Infrastructure IG**
  + **About**: ...to help repositories to preserve their data holdings...
  + **Intersections**: Semantic content of preserved data holdings.

# Work Plan

Work Plan: A specific and detailed description of how the WG will operate including: - Adoption Plan: A specific plan for adoption or implementation of the WG outcomes within the organizations and institutions represented by WG members, as well as plans for adoption more broadly within the community. Such adoption or implementation should start within the 12-18 month timeframe before the WG is complete. - The form and description of final deliverables of the WG, - The form and description of milestones and intermediate documents, code or other deliverables that will be developed during the course of the WG’s work, - A description of the WG’s mode and frequency of operation (e.g. on-line and/or on-site, how frequently will the group meet, etc.), - A description of how the WG plans to develop consensus, address conflicts, stay on track and within scope, and move forward during operation, and - A description of the WG’s planned approach to broader community engagement and participation.

## From our P8 Slides

### Timeline

* **Dec 16**: Identify initial set of focus areas and discuss.
* **Feb 17**: Draft case statements distributed.
* **Apr 17**: Discuss draft case statements and formation of WGs at Plenary 9.
* **May 17**: Finalize and circulate case statements for WGs.
* **May - Oct 17**: Short-term goals.
* **Sep 17**: Meeting WGs and summary of activity at Plenary 10.
* **Nov 17 - Apr 18**: Medium-term goals.
* **Apr 18**: Group health check Plenary 11.
* **May 18 - Oct 18**: Long-term goals.
* **Sep 17**: Final group Plenary 12.
* **October 18+**: After-term goals.

# Initial Membership

Initial Membership: A specific list of initial members of the WG and a description of initial leadership of the WG.

Membership will be sought from the Provenance IG and supplemented with a call to both other RDA groups and known non-RDA provenance communities, such as the provenance research community.

Since this group's work is likely to be highly relevant to, or even directed by, other WGs, it may be sensible to have other WG members attend this group's meetings either in a liaison role or as members in their own right.

1. First line mostly from the IG Charter [↑](#footnote-ref-23)
2. From our P8 slides, with the UC addition [↑](#footnote-ref-24)