

# Towards a Working Group Proposal

Nick Car and Dave Dubin

September 17, 2016

# Summary

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

- ▶ Provenance vocabularies offer a level of generality/specificity that address what we perceive to be implementation gaps.

# Summary

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

- ▶ 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.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.
- ▶ **December 2016:** Identify initial set of focus areas and discuss.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.
- ▶ **December 2016:** Identify initial set of focus areas and discuss.
- ▶ **February 2017:** Draft case statements distributed.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.
- ▶ **December 2016:** Identify initial set of focus areas and discuss.
- ▶ **February 2017:** Draft case statements distributed.
- ▶ **April 2017:** Discuss draft case statements and formation of WGs at Plenary 9.



## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.
- ▶ **December 2016:** Identify initial set of focus areas and discuss.
- ▶ **February 2017:** Draft case statements distributed.
- ▶ **April 2017:** Discuss draft case statements and formation of WGs at Plenary 9.
- ▶ **May 2017:** Finalize and circulate case statements for WGs.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.
- ▶ **December 2016:** Identify initial set of focus areas and discuss.
- ▶ **February 2017:** Draft case statements distributed.
- ▶ **April 2017:** Discuss draft case statements and formation of WGs at Plenary 9.
- ▶ **May 2017:** Finalize and circulate case statements for WGs.
- ▶ **May–September 2017:** Initial Working Group activity.

## Proposed timeline

- ▶ **Now:** Kickoff at RDA Plenary 8.
- ▶ **October 2016:** Establish web presence and mailing list, hold initial virtual meeting, prioritize deliverables, distribute workload.
- ▶ **December 2016:** Identify initial set of focus areas and discuss.
- ▶ **February 2017:** Draft case statements distributed.
- ▶ **April 2017:** Discuss draft case statements and formation of WGs at Plenary 9.
- ▶ **May 2017:** Finalize and circulate case statements for WGs.
- ▶ **May–September 2017:** Initial Working Group activity.
- ▶ **September 2017:** Meeting WGs and summary of activity at Plenary 10.

# Initial WG activity points

1. **Provenance patterns**

# Initial WG activity points

## 1. 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.

# Initial WG activity points

## 1. 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.

# Initial WG activity points

## 1. 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.

## 2. Sharing provenance

# Initial WG activity points

## 1. 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.

## 2. 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.



# Activity points

## 3. **Strategies for implementation**

# Activity points

## 3. 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.

# Activity points

## 3. **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.

## 4. **Tools for provenance**

# Activity points

## 3. **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.

## 4. **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).

## Other groups with which we likely have links

- ▶ **Publishing Data Workflows WG**



## Other groups with which we likely have links

- ▶ **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.”

## Other groups with which we likely have links

- ▶ **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:**

## Other groups with which we likely have links

- ▶ **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.



## Other groups with which we likely have links

- ▶ **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.

# Other groups with which we likely have links

## ▶ **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 <sup>1</sup>).

---

<sup>1</sup><http://www.scidatacon.org/2016/sessions/53/paper/47/>

# Other groups with which we likely have links

## ▶ **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 <sup>1</sup>).
  - ▶ Standards for data publishing ... and citation.

---

<sup>1</sup><http://www.scidatacon.org/2016/sessions/53/paper/47/>

## Other Groups

- ▶ **Dynamic Data Citation WG**

## Other Groups

- ▶ **Dynamic Data Citation WG**

- ▶ Describing the conceptual model containing object relevant to dynamic data citation.

## Other Groups

- ▶ **Dynamic Data Citation WG**
  - ▶ Describing the conceptual model containing object relevant to dynamic data citation.
- ▶ **PID Information Types**

## 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.

## 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:**



## 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.”

## 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**

## 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.

## 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.

## 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**

## 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

## 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**



## 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).

## 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.

## 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**

## 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?

## 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.

## 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**

## 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. . .

## 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.



Questions and suggestions?