Extending DCAM for Metadata Provenance

Kai Eckert
Mannheim University Library
Mannheim, Germany
eckert@bib.uni-mannheim.de

Daniel Garijo
Universidad Politécnica de Madrid
Madrid, Spain
dgarijo@delicias.dia.fi.upm.es

Michael Panzer
OCLC Online Computer Library Center, Inc.,
Dublin, Ohio, USA
panzerm@oclc.org

DC-2011
September 22nd 2011
The Hague, Netherlands
DCMI Metadata Provenance Task Group

- Members:
  - Kai Eckert (Mannheim University Library, Germany)
  - Daniel Garijo (Universidad Politécnica de Madrid, Spain)
  - Michael Panzer (OCLC, USA)
- Established: June 2010 (duration: 1 year, or so...)
- Goal: A Dublin Core conformant way to represent metadata provenance.
- Further Information:
  - http://www.dublincore.org/groups/provenance/
  - http://wiki.bib.uni-mannheim.de/dc-provenance/
Our motivation

- Back in 2009, we wanted Metametadata...
  - ... to add debugging information to metadata statements created with crosswalks,
  - ... to add source information to metadata statements merged from different sources,
  - ... to add additional information like confidence values to automatic indexing results, which is missing in ALL bibliographic metadata formats.
- Those were the days in Seoul... back in 2009.
Provenance

• Today, we talk about *provenance*.
  
  • “Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource.” W3C Provenance Incubator Group

• Provenance is everywhere and provenance can be described using **Dublin Core**.

• But there are many different ways to integrate the provenance information.
The Provenance of the Metadata

The Provenance of Mona Lisa

Leonardo di ser Piero DA VINCI, dit Léonard de Vinci

Musée du Louvre

Portait de Lisa Gherardini, épouse de Francesco del Giocondo, dite Mona Lisa, la Gioconda ou la Joconde
Description Set

Annotations

ex:book1 \(\xrightarrow{\text{dc:subject}}\) ex:annotation1

ex:someIndexingProcess

dc:creator

ex:hasConfidenceValue

ex:assignedSubject

"Some Topic"
UML Class Model

DCAM

Proposed DCPROV Extension
The Element Set

- What elements do we propose to use to describe the metadata provenance?
The Element Set

- What elements do we propose to use to describe the metadata provenance?
- Guess what:

Dublin Core
The Element Set

• What elements do we propose to use to describe the metadata provenance?

• Guess what:

Dublin Core

• We are discussing additional elements:
  • dcprov:sourceModified as subproperty of dc:source
  • dcprov:creationType, dcprov:rank
  • ...
RDF Implementation
Using Named Graphs

Description Set

Annotation Set
Newspaper travel guides example

- Travel guides from many different places around the world.
- 1st level provenance information regarding the guides: author, date of publication...
- 2nd level provenance information regarding the metadata of the guides: publisher, author, license...
- 3rd level provenance information about the RDF serialization displayed to the user: query executed, time of execution...
El Viajero travel guides example: Modeling

- 1<sup>st</sup> level: description set (metadata about a guide)
- 2<sup>nd</sup> level: annotation set (metadata provenance)
El Viajero travel guides example: Modeling (2)

- 3rd level: Another annotation set, describing the creation process of the data delivered to the user.
OAI-PMH example: Modeling
Discoverability

- Use of SPARQL to retrieve annotation and description sets

- Example:

```sparql
SELECT ?ds ?p ?o WHERE {
    GRAPH ?ds {
    }
    GRAPH ?as {
        ?as rdf:type dcprov:AnnotationSet .
    }
}
```
Provenance Workshop

- We have a half-day workshop tomorrow afternoon: **Friday, 2 PM, Room B/C**
- Open agenda:
  - Discuss questions regarding our work.
  - What is needed for an application profile?
  - Connection to W3C Provenance Working Group.
  - Connection to W3C RDF Working Group.
  - What is the semantic of a description set?
  - ...
Remaining steps

- We want to finish the task group shortly after DC-2011.
- Now is the time for feedback that can be incorporated into the application profile.
- Any help regarding the formulation of the application profile or other kinds of deliverables is highly appreciated.
- Please sign up to the low traffic mailing list: http://www.jiscmail.ac.uk/lists/dc-provenance.html
To summarize:

Make description sets identifiable.

Create descriptions of description sets.

And Dublin Core is ready for Metadata Provenance.

Thank you.