DCMI Metadata Provenance Task Group

Metadata Provenance
Dublin Core on the Next Level

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

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

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

Motivation

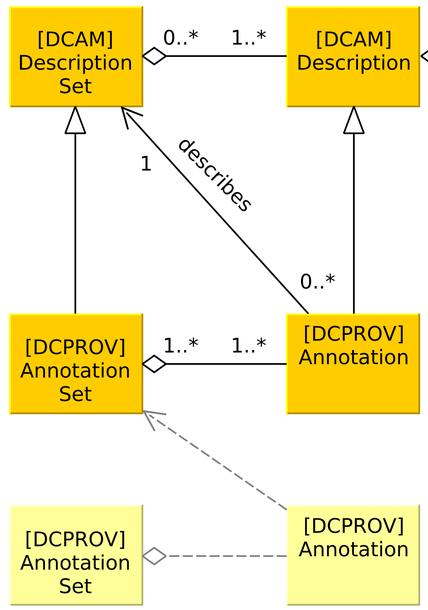
Our motivation for a Dublin Core application profile for metadata provenance is twofold:

Firstly, we want to **represent existing metadata provenance information** in a simple and unified way that is well suited as for an application of Dublin Core.
Secondly, we want to enable the provision of **provenance information for Dublin Core metadata** in a Dublin Core compatible way.

Element Set

In principle, the description of description sets is not different from the description of other resources. Therefore, we simply use **Dublin Core** as vocabulary on the provenance level, too.

As usual, the vocabulary can be extended and mixed with other vocabularies in application profiles. We are currently gathering special terms that might be useful to close gaps in Dublin Core.



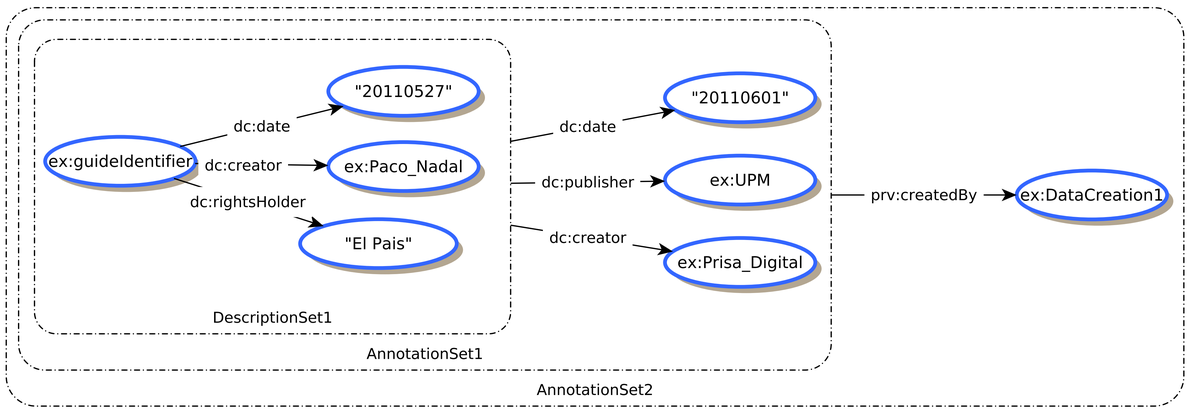
UML Class Model

We extend the Dublin Core Abstract Model (DCAM) with two new classes:

Description Set (from DCAM terminology): A set of one or more Descriptions, each of which describes a single resource.
Description (from DCAM terminology): One or more Statements about one, and only one, resource.
Statement (from DCAM terminology): An instantiation of a property-value pair made up of a property URI (a URI that identifies a property) and a value surrogate.
[NEW] Annotation: One or more Statements about one Description Set. Subclass of Description.
[NEW] Annotation Set: A set of one or more Annotations. Subclass of Description Set.

RDF Implementation using Named Graphs

The following **RDF graph** illustrates two levels of provenance information. The original **description set** describes a tourist guide. The provenance of the description data is stored in an **annotation set**. On the **second level**, the provenance for the actual data creation and delivery in a web application is stored.



Discoverability

Discovery is a two-stage process. First, description sets the triple is part of have to be determined, then the existence of an annotation set has to be determined for each. To find all provenance-related statements for some interpretation of a triple, the following SPARQL query can be used:

```
SELECT ?ds ?p ?o WHERE {
  GRAPH ?ds { ex:guideIdentifier dc:creator ex:Paco_Nadal . }
  GRAPH ?as { ?ds ?p ?o .
              ?as rdf:type dcprov:AnnotationSet . }
}
```

Further Information

http://dublincore.org/groups/provenance

DC-2011, September 21 - 23, 2011, The Hague, NL | Credits: Photography "Chain" by Zsuzsanna Kilian