

OPM Profile for Dublin Core Terms (Draft)

Simon Miles, Luc Moreau, Joe Futrelle

June 10, 2009

This document describes how provenance-related Dublin Core metadata terms [1] map to OPM graphs. The intention is to allow existing Dublin Core provenance to be re-expressed in OPM, so that it can be connected to wider provenance information available in OPM data.

This is an early, incomplete draft. Aside from checking the approach as a whole is valid, the following need to be completed.

- Provide proper graphs for OPM mappings (as in Figure 1).
- For some terms (available, valid), these can specify a single date or a period, but we currently only provide mapping for the former.
- Give a full worked example.
- Decide what an ‘actor’ is in OPM.

1 Prerequisite Concepts

Dublin Core terms are metadata about *resources*, which do not have a direct correspondence in OPM. Therefore, the mapping process must start by defining how resources and functionality which manipulates them, should be mapped to OPM.

1.1 Resources

An OPM artifact can represent a data or physical *resource* in one of several states which it takes over time. Therefore, we need a notion of one artifact being the *same resource as* another artifact.

The *same resource as* relation has a domain-dependent interpretation, but can be viewed as a sub-type of the OPM *was derived from* relation because one version of the resource will be a derivation from a previous version of the same resource. We give it the URI below, to help interoperability in URI-based instantiations of OPM:

`http://openprovenance.org/dublincore#sameResourceAs`

1.2 Modification

Resources change from state to state (artifact to artifact) by means of processes. Therefore, we define a process which takes a resource from one state to another as a *state-changing process* for that resource. In OPM, this is expressed as a process using one artifact as input and generating another artifact, with each artifact being the same resource as the other. Other artifacts may also be used or generated by the state-changing process.

A *creation process* for a resource is a process which generates an artifact corresponding to that resource, but for which there is no prior artifact (in the OPM graph) which is the same resource.

1.3 Actors

The values of Dublin Core metadata tags are often people, organisations, or services which have done something to the resources (created them, published them etc.) For brevity, we will call these kinds of entity *actors*, as, from a provenance perspective, we care about them only in as much as they act on resources/artifacts.

From an OPM perspective, the individual actions of actors are processes: they use and generate artifacts in specific instances. An actor itself, therefore, can be seen as an collection of processes (the set of those processes which represent its actions)¹.

We relate a process to the actor of which it is an action with the *was action of* relationship. We give it the URI below, to help interoperability in URI-based instantiations of OPM:

`http://openprovenance.org/dublincore#wasActionOf`

2 Dublin Core Term Mapping

2.1 Accrual Method

In Dublin Core, an *accrual method* is defined as “The method by which items are added to a collection” [1]. It is assigned the URL:

`http://purl.org/dc/terms/accrualMethod`

This term is not solely provenance-related: it may be about the past but is also about the future. We only consider it with regards to how it informs of how items *have been* added to a collection here.

In OPM, an *accrual method* to a collection resource is the actor performing a process to add items to that collection. Each instance of addition should be documented in the OPM graph, and the generalisation of those instances is the accrual method actor. For each addition, the following OPM pattern is

¹Is an actor equivalent to a process or an agent in OPM? If the former, is wasActionOf a sub-type of wasTriggeredBy?

observed, C being the accrual method, P being the process of addition of an item to a collection, A1 being the collection before adding item A2, A3 being the collection after adding A2:

```
A1 <-- used -- P
A2 <-- used -- P
P <-- wasGeneratedBy -- A3
A1 <-- sameResourceAs -- A3
C <-- wasActionOf -- P
```

Therefore to map a Dublin Core relationship A3 `dc:accrualMethod` C to OPM, we use the pattern above, adding an artifact to denote the resource before each addition took place (A1), the item added (A2), and the act of addition as a process (P).

More details about the way to model collection operations in OPM will be provided in a separate profile.

2.2 Available

In Dublin Core, *available* is defined as the “Date (often a range) that the resource became or will become available.” [1]. It is assigned the URL:

<http://purl.org/dc/terms/available>

In OPM, *being available* is part of the a state which a resource may take, and therefore corresponds to a subset of artifacts corresponding to that resource.

When referring to the future, the available date is simply an annotation to any artifact, marking when the resource it corresponds to will be available. When in the past, the available date may have a richer mapping as follows.

The *available date* of a resource is the timestamp annotated to the generation of the artifact representing that resource first in an available state. The following OPM pattern is observed, with A2 as the artifact made available, A1 as the artifact before being made available, P the process by which it was made available, and T the time at which this process completed:

```
A1 <-- used -- P
P <-- wasGeneratedBy [T] -- A2
A1 <-- sameResourceAs -- A2
```

where part of the state of A2 is that it is available.

Therefore to map a Dublin Core relationship A2 `dc:available` T to OPM, where A2 is in an available state, we use the pattern above, adding the act of being made available as a process (P) and the resource prior to availability (A1).

2.3 Bibliographic Citation

In Dublin Core, a *bibliographic citation* is defined as “A bibliographic reference for the resource” [1]. It is assigned the URL:

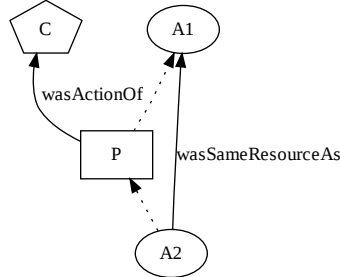


Figure 1: OPM graph mapping of a contributor term

<http://purl.org/dc/terms/bibliographicCitation>

As all bibliographic information may be contained in an OPM graph, including much of that mapped from other Dublin Core terms, the bibliographic citation can be seen as the results of a query over the OPM graph. This can return, for example, the creator, contributors, date published, method and collection in which it was published, all of which information should be present in the OPM graph.

When a generic query language for OPM has been expressed, this profile may be revised to define a query for retrieving a bibliographic citation for a resource.

2.4 Contributor

In Dublin Core, a *contributor* is defined as “An entity responsible for making contributions to the resource” with the comment “Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity” [1]. It is assigned the URL:

<http://purl.org/dc/terms/contributor>

In OPM, a *contributor* to a resource is the actor performing the state-changing process for that resource. The OPM pattern shown in Figure 1 is observed, with C being the contributor and A2 being the resource following contribution:

Therefore to map a Dublin Core relationship `A2 dc:contributor C` to OPM, we use the pattern in Figure 1, adding an artifact to denote the resource before the contribution took place (A1) and the act of contribution as a process (P).

2.5 Creator

In Dublin Core, a *contributor* is defined as “An entity primarily responsible for making the resource” with the comment “Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity” [1]. It is assigned the URL:

`http://purl.org/dc/terms/creator`

In OPM, the *creator* of a resource is the actor performing the creation process for that resource. A creation process is as defined above: a process which generates an artifact corresponding to that resource, where there should be no causally-prior artifact in the OPM graph which is the same resource as the created artifact. The following OPM pattern is observed, with C being the creator and A1 being the created resource:

```
P <-- wasGeneratedBy -- A1
C <-- wasActionOf -- P
THERE DOES NOT EXIST A0 SUCH THAT:
  A0 <-- sameResourceAs -- A1
```

Therefore to map a Dublin Core relationship A1 `dc:creator` C to OPM, we use the pattern above, adding the act of creation as a process (P) and ensuring no prior artifact of the same resource exists in the graph.

2.6 Date

In Dublin Core, a *date* is defined as “A point or period of time associated with an event in the lifecycle of the resource” with the comment “Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601” [1]. It is assigned the URL:

`http://purl.org/dc/terms/date`

Applying times to artifacts and processes is already covered by the core OPM specification.

2.7 Date Accepted

In Dublin Core, *date accepted* is defined as the “Date of acceptance of the resource.” with the comment “Examples of resources to which a Date Accepted may be relevant are a thesis (accepted by a university department) or an article (accepted by a journal).” [1]. It is assigned the URL:

`http://purl.org/dc/terms/dateAccepted`

The same mapping specified for *available* above applies here, but with the resource entering an ‘accepted’ instead of ‘available’ state.

2.8 Date Copyrighted

In Dublin Core, *date copyrighted* is defined as the “Date of copyright.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/dateCopyrighted`

The same mapping specified for *available* above applies here, but with the resource entering an ‘copyrighted’ instead of ‘available’ state.

2.9 Date Submitted

In Dublin Core, *date submitted* is defined as the “Date of submission of the resource.” with the comment “Examples of resources to which a Date Submitted may be relevant are a thesis (submitted to a university department) or an article (submitted to a journal)” [1]. It is assigned the URL:

`http://purl.org/dc/terms/dateSubmitted`

The same mapping specified for *available* above applies here, but with the resource entering an ‘submitted’ instead of ‘available’ state.

2.10 Has Part

In Dublin Core, *has part* is defined as “A related resource that is included either physically or logically in the described resource.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/hasPart`

In the OPM collections profile, the relation *contains* relates a collection to an item in that collection. Therefore to map a Dublin Core relationship `A1 dc:hasPart A2` to OPM, we say `A2 <-- oc:contains -- A1`, where `oc` is the OPM collections profile namespace.

2.11 Has Version

In Dublin Core, *has version* is defined as “A related resource that is a version, edition, or adaptation of the described resource.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/hasVersion`

In OPM, this corresponds to the *same resource as* relationship given above. Therefore, to map a Dublin Core relationship `A2 dc:hasVersion A1` to OPM, we use the pattern `A2 <-- odc:sameResourceAs -- A1` if `A1` is a later version than `A2` or `A1 <-- odc:sameResourceAs -- A2` if `A1` is earlier, where `odc` is the namespace used in this profile.

2.12 Identifier

In Dublin Core, *identifier* is defined as “An unambiguous reference to the resource within a given context.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/hasPart`

The definition of identifier corresponds to that of *artifact ID* in OPM: this unambiguously identifies a resource in a given context. Therefore, to map a Dublin Core relationship `A dc:identifier I` to OPM, we may use `I` as the artifact ID for the artifact, or resource in context, `A`. Where this is not appropriate, the relationship should be included as an annotation to the artifact.

2.13 Is Part Of

In Dublin Core, *is part of* is defined as “A related resource in which the described resource is physically or logically included.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/isPartOf`

In the OPM collections profile, the relation *contains* relates a collection to an item in that collection. Therefore, to map a Dublin Core relationship `A1 dc:isPartOf A2` to OPM, we say `A1 <-- oc:contains -- A2`, where `oc` is the OPM collections profile namespace.

2.14 Is Replaced By

In Dublin Core, *is replaced by* is defined as “A related resource that supplants, displaces, or supersedes the described resource.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/isReplacedBy`

In OPM, *replaces* is a sub-type of *was derived from*, denoting that one artifact is a replacement for another artifact. Therefore, to map a Dublin Core relationship `A1 dc:isReplacedBy A2` to OPM, we say `A1 <-- odc:replaces -- A2`, where `odc` is the namespace used in this profile.

2.15 Issued

In Dublin Core, *issued* is defined as the “Date of formal issuance (e.g., publication) of the resource.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/issued`

The same mapping specified for *available* above applies here, but with the resource entering an ‘issued’ instead of ‘available’ state.

2.16 Is Version Of

In Dublin Core, *is version of* is defined as “A related resource of which the described resource is a version, edition, or adaptation.” with the comment “Changes in version imply substantive changes in content rather than differences in format.” [1]. It is assigned the URL:

<http://purl.org/dc/terms/isVersionOf>

In OPM, *is version of* is a sub-type of *same resource as*, denoting that one artifact is a different version of the same resource represented as another artifact. Therefore, to map a Dublin Core relationship `A2 dc:isVersionOf A1` to OPM, where `A1` is an earlier version than `A2`, we say `A1 <-- odc:isVersionOf -- A2`, where `odc` is the namespace used in this profile.

2.17 Modified

In Dublin Core, *modified* is defined as the “Date on which the resource was changed.” [1]. It is assigned the URL:

<http://purl.org/dc/terms/modified>

A *modified date* of a resource is a timestamp annotated to the generation of any artifact representing that resource. The following OPM pattern is observed, with `A2` as the artifact after modification, `A1` as the artifact before modification, `P` the process by which it was modified, and `T` the time at which this process completed:

```
A1 <-- used -- P
P <-- wasGeneratedBy [T] -- A2
A1 <-- sameResourceAs -- A2
```

Therefore, to map a Dublin Core relationship `A2 dc:modified T` to OPM, we use the pattern above, adding the act of being modified as a process (`P`) and the resource prior to modification (`A1`).

2.18 Provenance

In Dublin Core, *provenance* is defined as “A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity, and interpretation.” [1]. It is assigned the URL:

<http://purl.org/dc/terms/provenance>

In OPM, the *provenance* of an artifact is an OPM graph in which there is a path from that artifact to every node (process, artifact) in the graph. Dublin Core-style provenance concerns the history of an artifact at a particular level of granularity, including only particular types of process, i.e. processes related to transfer of ownership.

Therefore, to map a Dublin Core relationship `A dc:provenance P` to OPM, we translate `P` to an OPM graph in which there is a path from `A` to every process, with these processes denoting the change in ownership.

2.19 Publisher

In Dublin Core, a *publisher* is defined as “An entity responsible for making the resource available” with the comment “Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity” [1]. It is assigned the URL:

`http://purl.org/dc/terms/publisher`

In OPM, *being published* is part of the a state which a resource may take, and therefore corresponds to a subset of artifacts corresponding to that resource.

The *publisher* of a resource is the actor performing the process which first generates an artifact which represents that resource and is in a published state. The following OPM pattern is observed, with C as the publisher, A2 as the artifact published, and A1 as the artifact before publication:

```
A1 <-- used -- P
P <-- wasGeneratedBy -- A2
A1 <-- sameResourceAs -- A2
C <-- wasActionOf -- P
```

where part of the state of A2 is that it is published.

Therefore to map a Dublin Core relationship **A2 dc:publisher C** to OPM, where A2 is in a published state, we use the pattern above, adding the act of publication as a process (P) and the resource prior to publication (A1).

2.20 Replaces

In Dublin Core, *replaces* is defined as “A related resource that is supplanted, displaced, or superseded by the described resource.” [1]. It is assigned the URL:

`http://purl.org/dc/terms/replaces`

In OPM, *replaces* is a sub-type of *was derived from*, denoting that one artifact is a replacement for another artifact. Therefore, to map a Dublin Core relationship **A2 dc:replaces A1** to OPM, we say **A1 <-- odc:replaces -- A2**, where odc is the namespace used in this profile.

2.21 Source

In Dublin Core, a *source* is defined as “A related resource from which the described resource is derived” with the comment “The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system” [1]. It is assigned the URL:

`http://purl.org/dc/terms/source`

In OPM, a *source* of an artifact is an artifact from which it is derived (explicitly or inferred). The following OPM pattern is observed:

```
A1 <-- wasDerivedFrom* -- A2
```

Therefore to map a Dublin Core relationship `A2 dc:source A1` to OPM, we use the pattern above.

2.22 Valid

In Dublin Core, *valid* is defined as the “Date (often a range) of validity of a resource.” [1]. It is assigned the URL:

```
http://purl.org/dc/terms/valid
```

In OPM, *being valid* is part of the a state which a resource may take, and therefore corresponds to a subset of artifacts corresponding to that resource.

When referring to the future, the valid date is simply an annotation to any artifact, marking when the resource it corresponds to will be valid. When in the past, the valid date may have a richer mapping as follows.

The *valid date* of a resource is the timestamp annotated to the generation of the artifact representing that resource first in a valid state. The following OPM pattern is observed, with A2 as the artifact made valid, A1 as the artifact before being made valid, P the process by which it was made available, and T the time at which this process completed:

```
A1 <-- used -- P
P  <-- wasGeneratedBy [T] -- A2
A1 <-- sameResourceAs -- A2
```

where part of the state of A2 is that it is available.

The same mapping specified for *available* above applies here, but with the resource entering an ‘valid’ instead of ‘available’ state.

3 Non-Provenance Dublin Core Terms

Some Dublin Core terms map to annotations on artifacts in OPM: abstract, accessRights, accrualPeriodicity, accrualPolicy, alternative, audience, conformsTo, coverage, description, educationLevel, extent, format, hasFormat, instructionalMethod, isFormatOf, isReferencedBy, isRequiredBy, language, license, mediator, medium, references, relation, rights, rightsHolder, spatial, subject, tableOfContents, title, type.

For a given artifact, an OPM query may be specified which searches for an annotation to any artifact which represents an earlier state of the same resource. In this way, annotations such as those above may be made to the resource in one context (one artifact) but be determinable given any artifact corresponding to that resource.

4 Vocabulary Encoding and Syntax Encoding Schemes

The Dublin Core vocabulary encoding schemes and syntax encoding schemes do not relate to provenance.

5 Classes and Type Vocabulary

Dublin Core specifies a set of classes and a type vocabulary, some of which have correspondence in OPM or this profile. The *Agent* class is equivalent to actors as defined in this profile. The *ProvenanceStatement* class corresponds to an OPM graph. Some classes describe sub-classes of resources, some are unions of other classes. However, none appear to have an impact on the mapping of Dublin Core data to OPM.

References

- [1] Dublin Core Metadata Initiative. Dublin core metadata terms. <http://dublincore.org/documents/dcmi-terms/>, May 2009.