

Open Archives Initiative Object Re-use & Exchange

Open Issues

Simeon Warner⁽¹⁾

Pete Johnston, Carl Lagoze, Michael Nelson,
Robert Sanderson, Herbert Van de Sompel

⁽¹⁾ Cornell Information Science

simeon@cs.cornell.edu

OAI-ORE Open Meeting
The Johns Hopkins University
Baltimore, MD, USA. 3 March 2008

Development process

pre-history Pathways work / Mellon meeting / Google meeting / TC meetings

alpha 0.1 10 December 2007

alpha 0.2 3 March 2008

beta April 2008

Significant change and evolution of ideas over the year before the alpha 0.1 release. Smaller changes 0.1 to 0.2 but some issues remain.

Extension element syntax in Atom

Q. Should the syntax and/or element-type of extension elements indicate/determine whether the object/content is a literal or a resource (URI)?

Obvious choice is to follow Atom style:

```
<atom:feed>
  ...
  <atom:author>
    <atom:name>Fred Bloggs</atom:name>
    <atom:uri>info:people/fred_bloggs</atom:uri>
  </atom:author>
  <ore:analogousTo>info:doi/1.23/abcde</ore:analogousTo>
  ...
</atom:feed>
```

Extension relationship syntax in Atom (2)

Should we instead have the following?

```
<atom:feed>
  ...
  <atom:author>
    <atom:name>Fred Bloggs</atom:name>
    <atom:uri>info:people/fred_bloggs</atom:uri>
  </atom:author>
  <ore:analogousTo ore:uri="info:doi/12.345/abcde"/>
  ...
</atom:feed>
```

“Better” but less in keeping with Atom.

Two serialization formats

At present we describe two serialization formats:

- **Atom**
- **RDF Syntaxes** and especially **RDF/XML**

Need to carefully describe appropriate choice (usually Atom) but also restrictions.

Q. Should we use a special name for the subset of the full ORE Model that Atom can serialize? Exactly how much should we include in the Atom serialization?

Choice of RDF/XML Profile

Q. Should the RDF/XML Profile say:

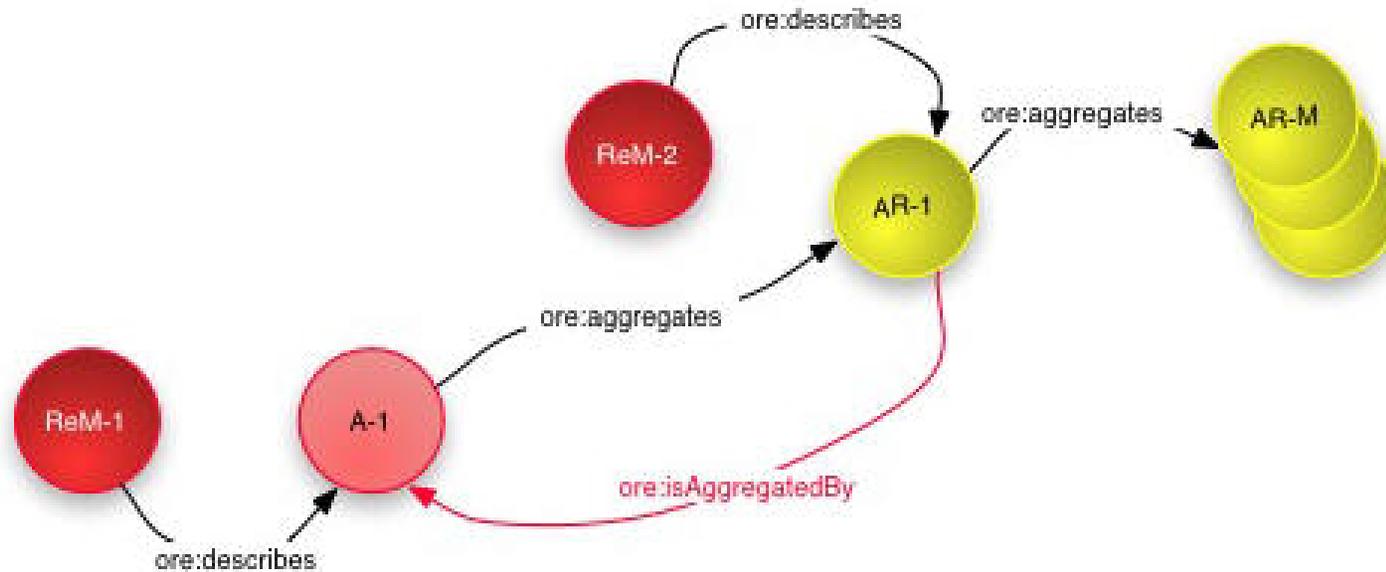
- Where an ORE Resource Map is serialized using RDF/XML, this profile of RDF/XML **MUST** be used; or
- Where an ORE Resource Map is serialized using RDF/XML, any of the conventions supported by the RDF/XML specification **MAY** be used; this profile is provided primarily as a guideline for Resource Map providers who are not familiar with the RDF/XML format.

Nested Aggregations

Let's examine something pretty much under control...

- Resource Maps describe only one level of aggregation
- If an Aggregated Resource is itself an aggregation then it must be described by a separate Resource Map.
- This second Resource Map may link back to the aggregating Resource Map with the `ore:isAggregatedBy` predicate (inverse of `ore:aggregates`).

Nested Aggregations (2)



<ReM-2>	<ore:describes>	<AR-1>
<AR-1>	<rdf:type>	<T-2>
<AR-1>	<ore:isAggregatedBy>	<ReM-1>

ore:isAggregatedBy	http://www.openarchives.org/ore/terms/isAggregatedBy
T-2	http://www.openarchives.org/ore/terms/Aggregation

[Abstract Data Model, section 5.7.2]

Reference in context

Citation is straightforward notion:

```
Resource1 dcterm:references Resource2.
```

Want the notion of “reference in context”: a reference to an Aggregated Resource *in the context of* an Aggregation.

Seemingly the only solution is to introduce an extra resource to stand for this contextualization:

```
Resource1 dcterm:references XXX.
```

```
XXX hasContext Aggregation2.
```

```
XXX hasReference Resource2.
```

Won't work with Atom, potential minting/maintenance issues for URI XXX, still difficult to define semantics well.

Alternatively require use of quad-store for context.

Lineage

Somewhat similar to reference in context but also about workflow: “I got AR-1 from A-1/ReM-1”.

We have explored use of a specialization of `ore:isAggregatedBy` that dumbs down to `ore:isAggregatedBy` but has additional meaning in context. However, this extends RDF notions by requiring context sensitive interpretation.

Some elements of these notions are already in Atom with the `/feed/entry/source` and `/feed/entry/link[@rel="via"]` elements. Not a perfect match.

Definition of ore:analogousTo

There is a critical need to associate the URIs of ORE Aggregations with other identifiers for aggregations. In the scholarly publishing world an important example is association with DOIs and legacy identifiers.

In almost all cases the relationship `owl:sameAs` is too tight – appropriate for same resource only.

Currently have `ore:analogousTo` as weaker relation (same intellectual object) but this specification needs to be tightened and spelled out more clearly.

Discovery of Resource Maps via HTML pages

Q. How can we provide a Resource Map context along with a link in an HTML page?

The following is nice but not valid HTML:

```
<a href="http://ex.org/pics/f-t.pdf"  
    resourcemap="http://ex.org/a.atom">frogs</a>
```

Might alternatively (mis)use the `class` attribute:

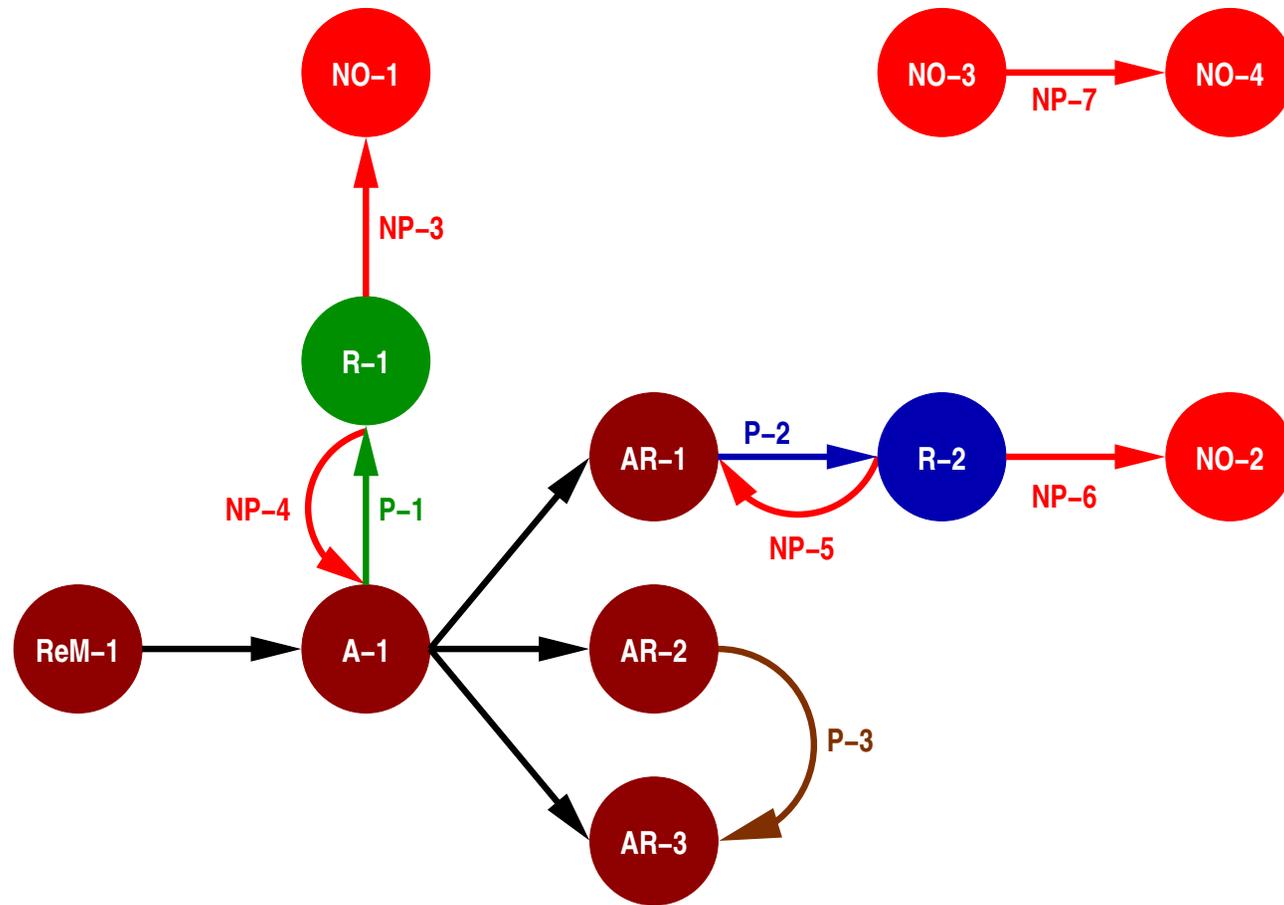
```
<a href="http://ex.org/pics/f-t.pdf"  
    class="resourcemap=http://ex.org/a.atom">frogs</a>
```

Uniform discovery of Resource Maps

We have discovery methods for external Resource Maps (HTML `<link>`, etc.).

RDFa permits embedded Resource Maps in HTML which could be detected by looking for them. However, it would be appealing to have a common pointing approach to simplify detection.

What relationships should be allowed?



No **NO** nodes or **NP** predicates permitted in Atom.

Disconnected component forbidden in ORE Model.

Choice regarding external vocabularies

Q. Should the ORE select or promote certain vocabularies as preferred?

To date we have created a minimal number of additional elements in the ORE namespace and adopted elements from Dublin Core, OWL and RDF. Should we propose or recommend additional elements or leave choice open to implementers?

Aggregation without enumeration

In current specifications all aggregated resources must be explicitly enumerated (each URI written in Resource Map).

There is considerable interest in extending the ORE Data Model to handle the notion of aggregation without explicitly enumerating all of the Aggregated Resources. Two cases:

- URI Templates — Obvious overlap with the **W3C POWDER Working Group** although the focus is quite different.
- Dynamically generated — or virtual collections. Useful to be able to describe without actually generating or in a way constant over differing generations.