Open Archives Initiative Object Re-use & Exchange
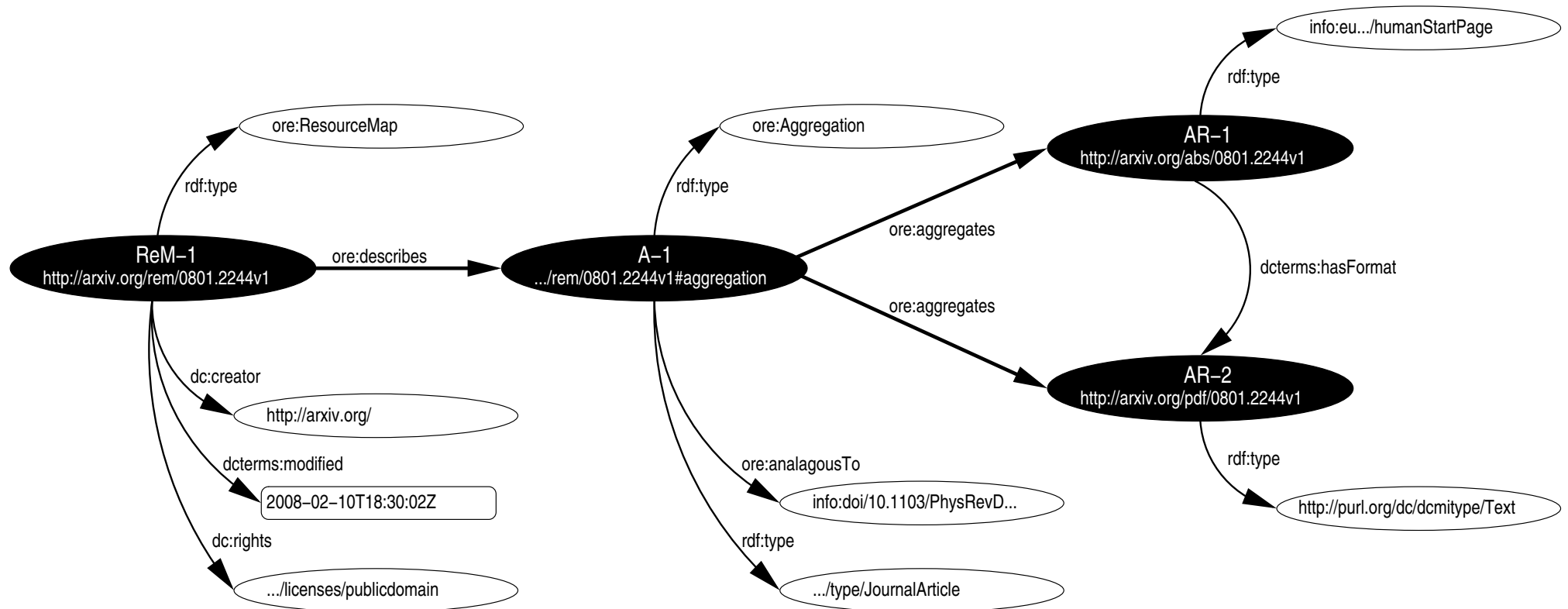
Serialization in Atom

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

# Let's build an Atom Resource Map

Very simple case of an article on arXiv for which we build an aggregation comprising the splash page and a PDF.

Resource Map, Aggregation and Aggregated Resources

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom"> <!-- Resource Map -->
  <id>tag:arxiv.org,2008:0801.2244v1</id>
  <link rel="describes"
        href="http://arxiv.org/rem/0801.2244v1#aggregation" />

  <entry> <!-- AR-1 -->
    <id>tag:arxiv.org,2008:0801.2244v1:abs</id>
  </entry>
  <entry> <!-- AR-2 -->
    <id>tag:arxiv.org,2008:0801.2244v1:pdf</id>
  </entry>
</feed>
```
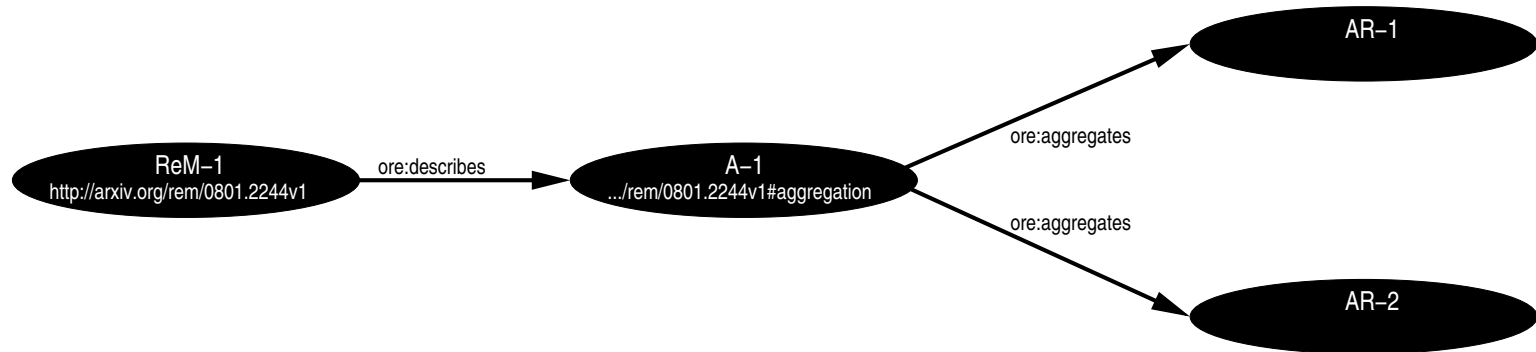
Structure BUT haven't yet identified aggregated resources.

Atom identifiers

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom"> <!-- Resource Map -->
  <id>tag:arxiv.org,2008:0801.2244v1</id>
  <link rel="describes"
        href="http://arxiv.org/rem/0801.2244v1#aggregation" />

  <entry> <!-- AR-1 -->
    <id>tag:arxiv.org,2008:0801.2244v1:abs</id>
  </entry>
  <entry> <!-- AR-2 -->
    <id>tag:arxiv.org,2008:0801.2244v1:pdf</id>
  </entry>
</feed>
```

Atom `<id>` elements have no ORE meaning.

The `tag` or `uuid` URI schemes can provide convenient (no-cost) way to provide for Atom.

Typing and Resource Map URI

```xml
<feed xmlns="http://www.w3.org/2005/Atom">
  <id>tag:arxiv.org,2008:0801.2244v1</id>
  <link href="http://arxiv.org/rem/0801.224v1" rel="self" type="application/atom+xml"/>
  <category scheme="http://www.openarchives.org/ore/terms/"
            term="http://www.openarchives.org/ore/terms/ResourceMap" label="Resource Map" />
  <link rel="describes" href="http://arxiv.org/rem/astro-ph/0601007#aggregation" />

  <entry> <!-- AR-1 -->
    <id>tag:arxiv.org,2008:0801.2244v1:abs</id>
  </entry>
  <entry> <!-- AR-2 -->
    <id>tag:arxiv.org,2008:0801.2244v1:pdf</id>
  </entry>
</feed>
```

Resource Map URI in `/feed/link[@rel="self"]`
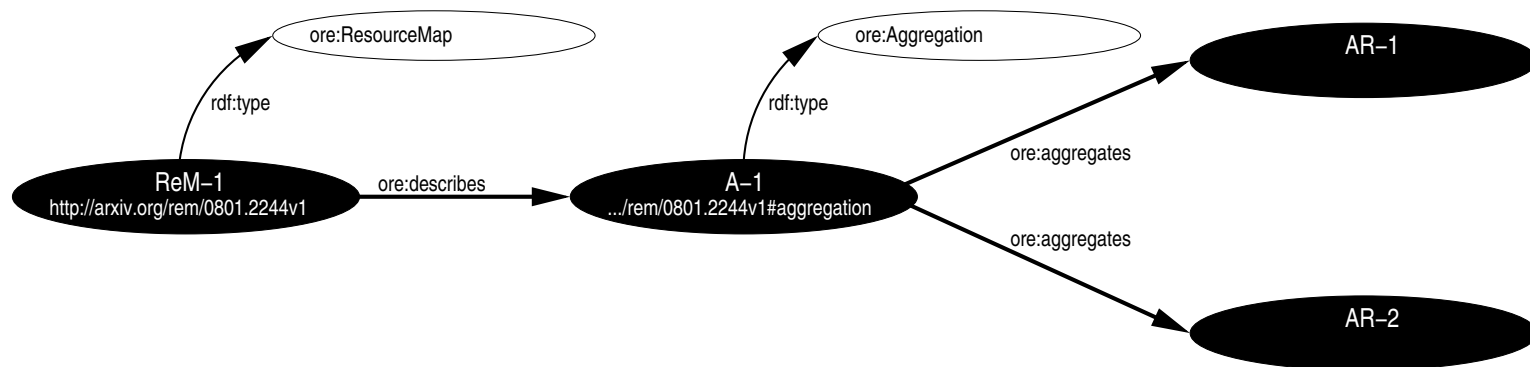
Aggregated Resource URIs

```xml
<feed xmlns="http://www.w3.org/2005/Atom">
  <id>tag:arxiv.org,2008:0801.2244v1</id>
  <link href="http://arxiv.org/rem/0801.2244v1" rel="self" type="application/atom+xml"/>
  <category scheme="http://www.openarchives.org/ore/terms/"
            term="http://www.openarchives.org/ore/terms/ResourceMap" label="Resource Map" />
  <link rel="describes" href="http://arxiv.org/rem/0801.2244v1#aggregation" />

  <entry>
    <id>tag:arxiv.org,2008:0801.2244v1:abs</id>
    <link href="http://arxiv.org/abs/0801.2244v1" rel="alternate" type="text/html"/>
 </entry>


 <entry>
    <id>tag:arxiv.org,2008:0801.2244v1:pdf</id>
    <link href="http://arxiv.org/pdf/0801.2244v1" rel="alternate" type="application/pdf"/>
 </entry>
</feed>
```
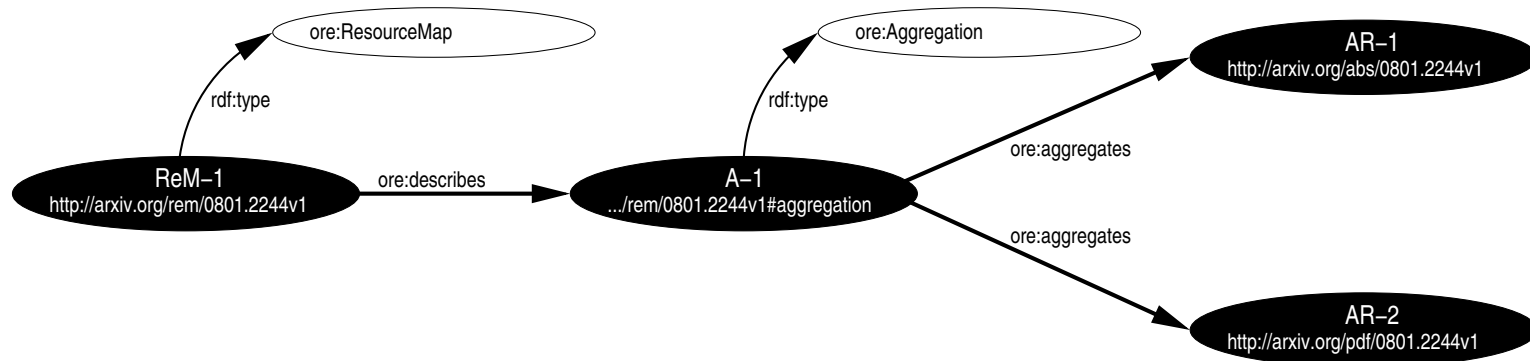
Atom feed and entry titles

```
<feed xmlns="http://www.w3.org/2005/Atom">
  <id>tag:arxiv.org,2008:0801.2244v1</id>
  <link href="http://arxiv.org/rem/0801.2244v1" rel="self" type="application/atom+xml"/>
  <category scheme="http://www.openarchives.org/ore/terms/"
            term="http://www.openarchives.org/ore/terms/ResourceMap" label="Resource Map" />
  <link rel="describes" href="http://arxiv.org/rem/0801.2244v1#aggregation" />
  <title>Resource Map http://arxiv.org/rem/0801.2244v1</title>

 <entry>
   <id>tag:arxiv.org,2008:0801.2244v1:ps</id>
   <link href="http://arxiv.org/abs/0801.2244v1" rel="alternate" type="text/html"/>
   <title>Aggregated Resource http://arxiv.org/abs/0801.2244v1</title>
 </entry>

 <entry>
   <id>tag:arxiv.org,2008:0801.2244v1:pdf</id>
   <link href="http://arxiv.org/pdf/0801.2244v1" rel="alternate" type="application/pdf"/>
   <title>Aggregated Resource http://arxiv.org/pdf/0801.2244v1</title>
 </entry>

</feed>
```

Not the titles of resources referred to. No ORE meaning.

Resource map creator

```xml
<feed xmlns="http://www.w3.org/2005/Atom">
  <id>tag:arxiv.org,2008:0801.2244v1</id>
  <link href="http://arxiv.org/rem/0801.2244v1" rel="self" type="application/atom+xml"/>
  <category scheme="http://www.openarchives.org/ore/terms/"
            term="http://www.openarchives.org/ore/terms/ResourceMap" label="Resource Map" />
  <link rel="describes" href="http://arxiv.org/rem/0801.2244v1#aggregation" />
  <title>Resource Map http://arxiv.org/rem/0801.2244v1</title>
  <author>
    <name>arXiv.org e-Print Repository</name>
    <uri>http://arxiv.org/</uri>
  </author>
      ...
</feed>
```

Again, *not* related to authorship of aggregated resources.
Use of URI *recommended*.

Update timestamps

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
      ...
   <updated>2008-02-10T18:30:02Z</updated>

   <entry>
      <id>tag:arxiv.org,2008:0801.2244v1:ps</id>
      ...
      <updated>2008-02-10T18:30:02Z</updated>
   </entry>

   <entry>
      <id>tag:arxiv.org,2008:0801.2244v1:pdf</id>
      ...
      <updated>2008-01-31T12:52:00Z</updated>
   </entry>
</feed>
```
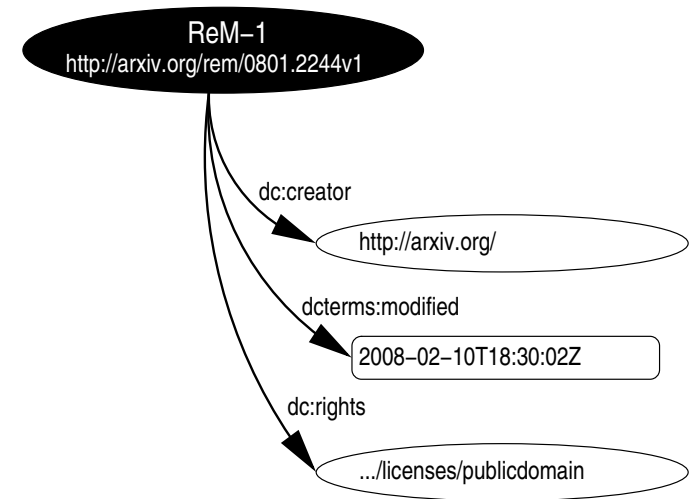
Resource Map (feed) timestamp required by model and Atom.
Aggregated Resource (entry) timestamps required by Atom.

Optional Resource Map Metadata

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
      ...
  <author>
    <name>arXiv.org e-Print Repository</name>
    <uri>http://arxiv.org/</uri>
  </author>
  <updated>2008-02-10T18:30:02Z</updated>
  <rights>http://creativecommons.org/public_domain<rights>
      ...
</feed>
```

Rights information (URI *strongly recommended*).

Relationships about the Aggregation

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
      ...
  <rights>http://creativecommons.org/public_domain</rights>
  <atom:link rel="related" href="info:doi/10.1103/PhysRevD10p1734"/>
  <rdf:type>info:eu-repo/semantics/JournalArticle</rdf:type>
  <entry>
      ...
  </entry>
</feed>
```

RULE: All elements included in the feed in the atom namespace except `/feed/link[@rel="related"]` are about the Resource Map, all other elements in the feed are about the Aggregation. Here we relate to a DOI and add a semantic type.
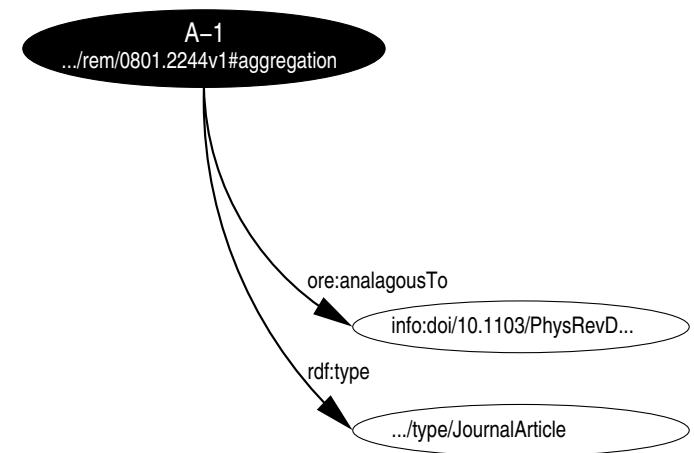
Type information about Aggregated Resources

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed xmlns="http://www.w3.org/2005/Atom">
      ...
  <entry>
    <id>tag:arxiv.org,2008:0801.2244v1:ps</id>
      ...
    <rdf:type>info:eu-repo/semantics/humanStartPage</rdf:type>
    <dcterms:hasFormat>http://arxiv.org/pdf/astro-ph/0601007v1</dcterms:hasFormat>
  </entry>
  <entry>
      ...
    <rdf:type>http://purl.org/dc/dcmitype/Text</rdf:type>
  </entry>
</feed>
```
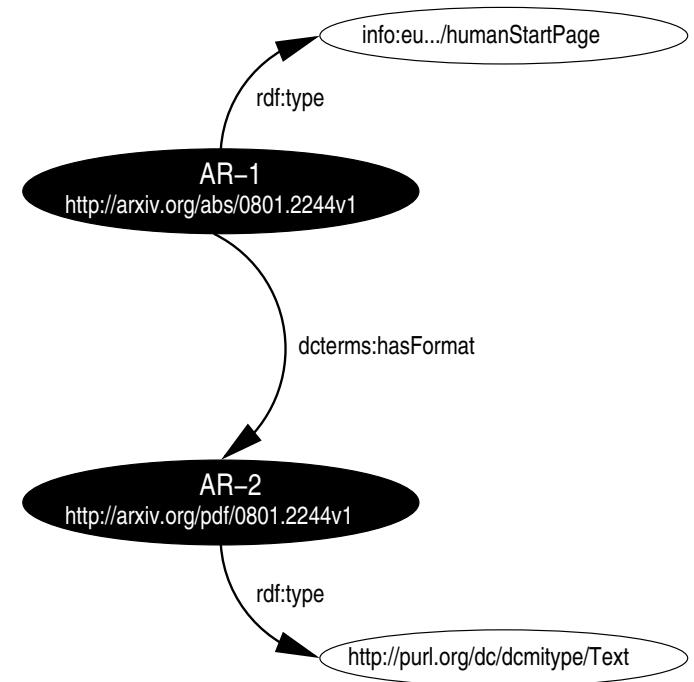
We have an HTML splash page and a PDF file.

Related with `dcterms:hasFormat`.

Not described

- Links to other aggregations — use of `/feed/entry/source` and `/feed/entry/link[@rel="via"]/@href` is described in implementation guide.

- Bibliographic metadata – one can include additional relationship for the Aggregation but likely best to aggregate metadata as another resource (e.g. serve via OAI-PMH).

- Common scenarios – implementation guide comments on mirroring, splash pages, versioning and multiple formats.

Tools

- **Atom Feed validator** — general purpose validator for Atom feed documents at `http://validator.w3.org/feed/check.cgi` (check ex4.2)
  Libraries available to automate use (e.g. `WebService::Validator::Feed::W3C` in Perl from CPAN)

- **Validator for ORE Resource Maps in Atom** — alpha version available at `http://www.openarchives.org/ore/atom-validator`.

Tools (2)

- **Schematron Schema for the Resource Map Profile** — available at `http://www.openarchives.org/ore/atom-tron`.

- **GRDDL crosswalk from Atom XML to RDF/XML** — available at `http://www.openarchives.org/ore/atom-grddl`.

Programmatic generation — Perl example

```perl
#!/usr/bin/perl
use lib '.';
use Net::OAI::ORE::ReM;

my $rem=Net::OAI::ORE::ReM->new(uri=>'http://arxiv.org/rem/0801.2244v1');
$rem->aggregated_resources('http://arxiv.org/abs/0801.2244v1',
                           'http://arxiv.org/pdf/0801.2244v1');
$rem->creator_uri('http://arxiv.org/');
$rem->modified_as_iso8601('2008-02-10T18:30:02Z');

print $rem->serialize('atom');
```

Programmatic generation — Perl example output

```xml
<?xml version="1.0" encoding="utf-8"?>
<feed ...  >
 <id>urn:uuid:409240A6-C2DF-3EB6-A4AA-48A5380B0746</id>
 <title>Resource Map http://arxiv.org/rem/0801.2244v1</title>
 <link rel="self" type="application/atom+xml" href="http://arxiv.org/rem/0801.2244v1" />
 <category term="http://www.openarchives.org/ore/terms/ResourceMap"
    label="Resource map" scheme="http://www.openarchives.org/ore/terms/" />
 <link rel="describes" href="http://arxiv.org/rem/0801.2244v1#aggregation" />
 <author>
  <uri>http://arxiv.org/</uri>
  <name>Authority http://arxiv.org/</name>
 </author>
 <updated>2008-02-10T18:30:02Z</updated>
 <dcterms:created>2008-02-10T18:30:02Z</dcterms:created>
 <entry>
  <id>urn:uuid:DCB09660-CD33-3BBB-90BD-51234F2663FB</id>
  <link rel="alternate" href="http://arxiv.org/abs/0801.2244v1" />
  <title>Aggregated Resouce http://arxiv.org/abs/0801.2244v1</title>
  <content>Aggregated Resouce http://arxiv.org/abs/0801.2244v1</content>
  <updated>2008-02-10T18:30:02Z</updated>
 </entry>
      ...
```

Atom vs RDF Syntaxes

pro Atom

- Sufficiently expressive for most purposes

- Widely used and understood

- Good tools and libraries

- RDF/XML can be generated via

pro RDF Syntaxes

- Complete expressive power

- No mapping required

- Easy extensibility

We've talked only about Atom so far, and that will be sufficient and easiest for most uses.

Types of relationship possible in Atom
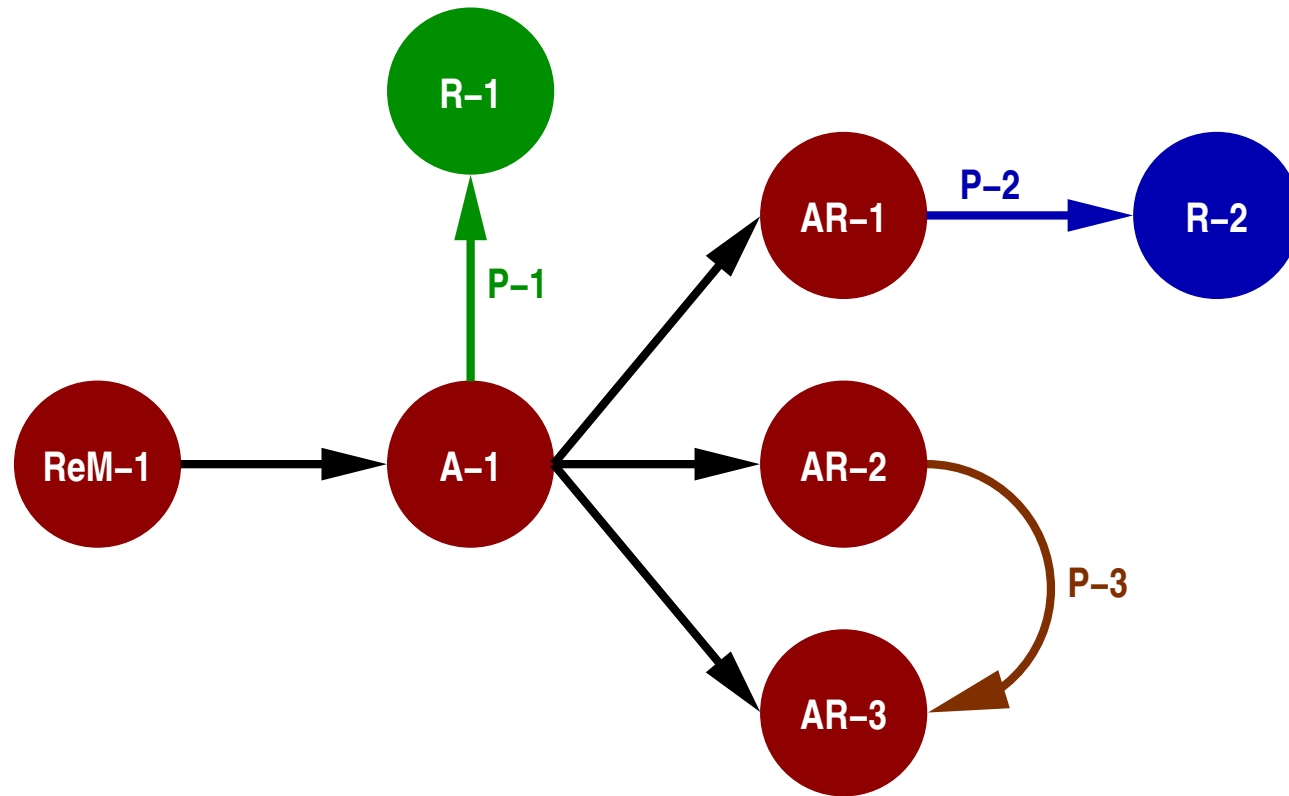
Where to start

- ORE User Guide - Overview for summary of data model.

- ORE User Guide - Resource Map Implementation in Atom for most of what you need to know to create Resource Maps.

- ORE Specification - Resource Map Profile of Atom for the gory details. Use as reference.

- Validators already described.

- OAI-ORE Google Group (`oai-ore`) for comments and discussion. We'd love additional feedback in this alpha phase and on through beta.

That's all folks...