

Open Archives Initiative Object Reuse & Exchange

Resource Map Discovery

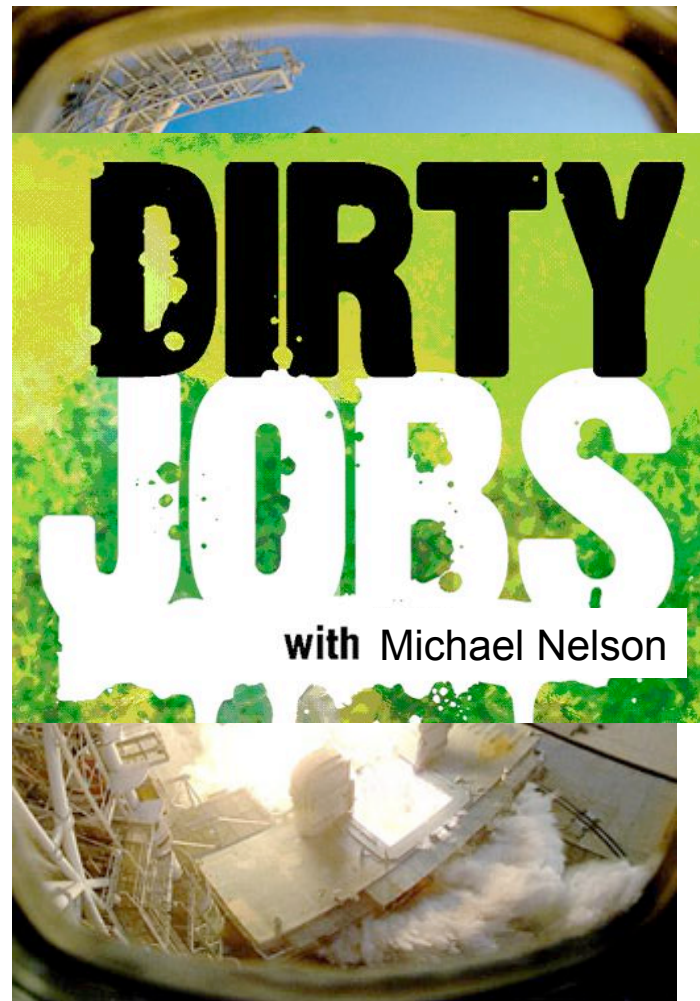
Michael L. Nelson*

Carl Lagoze, Herbert Van de Sompel, Pete Johnston, Robert Sanderson, Simeon Warner

OAI-ORE Specification Roll-Out
Baltimore MD, March 3, 2008

*Old Dominion University, Norfolk VA
<http://www.cs.odu.edu/~mln/>

Discovery...



Resource Map Discovery Outline

- Batch
 - OAI-PMH, SiteMaps, RSS/Atom
- Embedding
 - *ReMs in HTML (open issues)*
 - ReMs in non-HTML
- How not to do it
 - ReMs are not for humans
 - *URI conflation (open issues)*

Batch Discovery

- ReMs are resources and we already know how to expose large batches of resources:
 - OAI-PMH
 - SiteMaps
 - RSS/Atom

Batch :: ReMs in OAI-PMH

http://www.foo.edu/oai?verb=ListRecords&metadataPrefix=oai_rem

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2007-02-08T08:55:46Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_rem">
    http://foo.edu/oai2</request>
  <ListRecords>
    <record>
      <header>
        <identifier>oai:foo.edu:object1</identifier>
        <datestamp>2007-01-06</datestamp>
      </header>
      <metadata>
        <!-- Insert object1 ReM here -->
      </metadata>
    </record>
    . . .
  </ListRecords>
</OAI-PMH>
```

MUST NOT
equal either ReM Atom /feed/id or
/feed/link[@rel="self"]/@href

MUST be equal to ReM Atom /feed/updated

OAI-PMH GetRecord Processing

http://www.foo.edu/oai?verb=GetRecord&identifier=oai:foo.edu:object1&metadataPrefix=oai_rem

http://some.gateway.org/pmh2ore?=http://foo.edu/oai2?verb=GetRecord&metadataPefix=oai_rem&identifier=oai:foo.edu:object1

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2007-02-08T08:55:46Z</responseDate>
  <request verb="GetRecord" identifier="oai:foo.edu:object1"
    metadataPrefix="oai_rem">http://foo.edu/oai2</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:foo.edu:object1</identifier>
        <datestamp>2007-01-06</datestamp>
      </header>
      <metadata>
        <!-- Insert Object1 ReM here -->
      </metadata>
    </record>
  </GetRecord>
</OAI-PMH>
```

need a gateway to:

1. strip off OAI-PMH wrappers
2. return just what is inside <metadata>
3. reset the MIME type (e.g., from application/xml to application/atom+xml)

Batch :: ReMs in SiteMaps

http://www.foo.edu/sitemap-rem.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.foo.edu/objects/object1.atom</loc>
    <lastmod>2007-01-06</lastmod>
  </url>
  <url>
    <loc>http://www.foo.edu/objects/object2.atom</loc>
    <lastmod>2007-08-11</lastmod>
    <changefreq>weekly</changefreq>
  </url>
  <url>
    <loc>http://www.foo.edu/objects/object3.atom</loc>
    <lastmod>2007-03-15T18:30:02Z</lastmod>
    <priority>0.3</priority>
  </url>
  ...
</urlset>
```

MUST equal /feed/link[@rel="self"]/@href
for corresponding ReM, but
MUST NOT equal /feed/id

MUST be equal to ReM Atom /feed/updated

remember SiteMap path limitation: http://www.foo.edu/a/b/sitemap-rem.xml can list
http://www.foo.edu/a/b/bar2.atom but not http://www.foo.edu/bar1.atom

Batch :: ReMs in RSS

<http://www.foo.edu/all-rems.rss>

```
<?xml version="1.0"?>
<rss version="2.0">
  <channel>
    <title>ReMs at www.foo.edu</title>
    <link>http://www.foo.edu/</link>
    <description>All of the Resource Maps for resources at www.foo.edu</description>

    <item>
      <title>ReM for Object 1</title>
      <link>http://www.foo.org/objects/object1.atom</link>
      <description>ReM for Object 1</description>
      <pubDate>Sat, 06 Jan 2007 00:00:00 GMT</pubDate>
    </item>

    <item>
      <title>ReM for Object 2</title>
      <link>http://www.foo.org/objects/object2.atom</link>
      <description>ReM for Object 2</description>
      <pubDate>Sat, 11 Aug 2007 00:00:00 GMT</pubDate>
    </item>
  </channel>
</rss>
```

MUST NOT equal ReM Atom /feed/id;
MUST equal ReM Atom
/feed/link[@rel="self"]/@href

MUST equal ReM Atom /feed/updated
(after conversion from RFC-822 format to ISO 8601 format)

Batch :: ReMs in Atom

http://www.foo.edu/all-rems.atom

```
<feed xmlns="http://www.w3.org/2005/Atom">
  <title>ReMs at www.foo.edu</title>
  <link href="http://www.foo.edu/" />
  <link href="http://www.foo.edu/all-rems.atom" rel="self"/>
  <updated>2007-08-15T18:30:02Z</updated>
  <author>
    <name>John Doe</name>
    <email>johndoe@foo.edu</email>
  </author>
  <id>urn:uuid:60a76c80-d399-11d9-b91c-0003939e0af6</id>

  <entry>
    <title>ReM For Object1</title>
    <link href="http://www.foo.org/objects/object1.atom"/>
    <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
    <updated>2007-01-06T00:00:00Z</updated>
  </entry>

  <entry>
    <title>ReM For Object2</title>
    <link href="http://www.foo.org/objects/object2.atom"/>
    <id>urn:uuid:9a2cc699-ccba-9e8b-132e-91da394e9a5c</id>
    <updated>2007-08-11T00:00:00Z</updated>
  </entry>
</feed>
```

MUST equal ReM Atom
/feed/link[@rel="self"]/@href

MUST NOT equal ReM Atom /feed/id;

MUST equal ReM Atom /feed/updated

Embedding ReMs into Resources

- Starting with a resource, how to find the associated ReM(s)?
 - HTML `<link>`
 - HTTP `<A>` & ``
 - HTTP Response Headers
 - ReM Transparency
- 4 levels to describe resources' knowledge of their ReMs

Embedding :: Knowledge Levels

- Full knowledge
 - the ReM is linked to by all resources in the aggregation.
- Indirect knowledge
 - all but one of the resources in the aggregation link to a single, unique resource in the aggregation, which in turn links to the ReM.
 - functionally the same as full knowledge, but likely to be useful in actual deployment
- Limited knowledge
 - only a subset of the resources in the aggregation (typically just a single resource) link to the ReM, and the remainder of the resources have no links at all.
- Zero knowledge
 - none of the resources in the aggregation link to a ReM.

HTML `<link>` :: Full Knowledge

```
<html>
<head>
<title>Hello World.</title>
<link href="http://example.net/hw.atom"
      type="application/atom+xml" rel="resourceMap" >
</head>
<body>


</html>
```

HTML `<link>` :: Indirect Knowledge

```
<html>
<head>
<title>Chapter Twelve.</title>
<link href="http://mybook.com/toc.html"
      type="text/html" rel="indirectresource" >
</head>
<body>
Welcome to chapter twelve...
</body>
</html>
```

HTML `<link>` vs. `<A>` & ``

- `link` is from “this” document to its 1 or more corresponding ReMs
- `A` & `IMG` capabilities are proposed to provide “hints” about the context of the disaggregated resources
 - problem: HTML does not support statements of the form “I got this from there”
 - example: “I got this JPEG from ReM1, the PDF from ReM2 and this quoted text section from ReM3.”

HTML Option #1: `resourcemap` attribute

```
<html>
...
Here is a helpful reference for distinguishing
<a href="http://example.org/pics/f-t.pdf"
resourcemap="http://example.org/amphibians.atom">frogs vs. toads</a>.
<p>
Here is a frog

and here is a toad .
...
</html>
```

Pro: very simple, human readable

Con: invalid HTML

HTML Option #2:

<A> rel attribute

```
<html>
...
Here is a helpful reference for distinguishing
<a href="http://example.org/pics/f-t.pdf"
rel="resourcemap=http://example.org/amphibians.atom">frogs vs. toads</a>.
<p>
Here is a frog
<a rel="resourcemap=http://frogs.org/frogs.atom">

</a> and here is a toad
<a rel="resourcemap=http://toadsrule.org/toads.atom">

</a>.
...
</html>
```

Pro: Valid HTML

Con: Not uniform (<A> and do not (yet) support the same elements)

HTML Option #3: elements

```
<html>
...
Here is a helpful reference for distinguishing
<span class="resourcemap=http://example.org/amphibians.atom">
<a href="http://example.org/pics/f-t.pdf" frogs vs. toads</a>.
</span>
<p>
Here is a frog
<span class="resourcemap=http://frogs.org/frogs.atom">

</span> and here is a toad
<span class="resourcemap=http://toadsrule.org/toads.atom">

</span>.
...
</html>
```

Pro: Valid HTML, Uniform Approach

Con: No longer simple?

HTML Option #4: class attribute

```
<html>
...
Here is a helpful reference for distinguishing
<a href="http://example.org/pics/f-t.pdf"
class="resourcemap=http://example.org/amphibians.atom">frogs vs. toads</a>.
<p>
Here is a frog

and here is a toad .
...
</html>
```

Pro: very simple, human readable, valid HTML

Con: stretches, but does not break, "class"*

* <http://www.w3.org/TR/REC-html40/struct/global.html#edef-class>

The class attribute has several roles in HTML:

* As a style sheet selector (when an author wishes to assign style information to a set of elements).

* For general purpose processing by user agents.

Embedding :: ReM Transparency

- There is precedent for exposing URIs, JavaScript, etc. as opaque strings to users to paste into other applications
- This is not the same as creating a hypertext link to the scripts...

Embedding :: ReM Transparency

The screenshot shows a YouTube video player for the song "You Were Always The One" by The Crips. The video is currently at 01:40 of a 02:47 duration. The channel is "wchitarecordings", which has 47 videos and was joined 1 year ago. The video has 36,551 views and a rating of 4.5 stars from 103 ratings. The embed code is highlighted in a red box:

```
<object width="425" height="355"><param name="movie" value="h
```

Below the video player, there are links for Share, Favorite, Add to Playlists, and Flag. The video player also includes a progress bar, volume control, and a play/pause button.

Embedding :: ReM Transparency

rmccartney/1955 Nomad - Photobucket - Video and Image Hosting









http://s5.photobucket.com/albums/y167/rmccartney/1955%20Nomad

photobucket help | login

home | join now | find stuff | images | videos | web Search

Album: [rmccartney](#) > **1955 Nomad** (All, 30 images) links grid

1 2 next | view all

			
Email & IM http://s5.photobuc	Email & IM http://s5.photobuc	Email & IM http://s5.photobuc	Email & IM http://s5.photobuc
Direct Link http://i5.photobucl	Direct Link http://i5.photobucl	Direct Link http://i5.photobucl	Direct Link http://i5.photobucl
HTML Code <code><img src="http://i5</code>	HTML Code <code><img src="http://i5</code>	HTML Code <code><img src="http://i5</code>	HTML Code <code><img src="http://i5</code>
IMG Code <code>[IMG]http://i5.phot</code>	IMG Code <code>[IMG]http://i5.phot</code>	IMG Code <code>[IMG]http://i5.phot</code>	IMG Code <code>[IMG]http://i5.phot</code>
			
Email & IM http://s5.photobuc	Email & IM http://s5.photobuc	Email & IM http://s5.photobuc	Email & IM http://s5.photobuc
Direct Link http://i5.photobucl	Direct Link http://i5.photobucl	Direct Link http://i5.photobucl	Direct Link http://i5.photobucl
HTML Code <code><img src="http://i5</code>	HTML Code <code><img src="http://i5</code>	HTML Code <code><img src="http://i5</code>	HTML Code <code><img src="http://i5</code>

Find: Next Previous Highlight all

Done

Embedding :: ReM Transparency

Hemmings Motor News: Auto Classifieds - Hemmings Auto Classifieds feature cars for sale nation wide.

http://www.hemmings.com/classifieds/carsforsale/ford/fairlane/?y

Home Classifieds My Hemmings Publications Research Store Directory Forums Downloads About Us

Search All Classifieds Members: [Log in](#) | New User? [Register](#)

Classifieds > Cars For Sale > Ford > Fairlane

(VIEWING YEAR 1966)


Placing your ad is easy and affordable. Ads start at only \$11.70 for 18 words. Subscribers pay only \$9.90. [Click to learn more.](#)

You are browsing ads that are from the February 2008 issue of Hemmings Motor News. To see the most current ads from our print edition, you need to be a subscriber to Hemmings Motor News. If you are already a subscriber [link your subscription](#). If you would like to subscribe [Click here](#).

Viewing Ads 1 - 2 of 2
1 |
Sort: [Date Posted](#) | [City](#) | [Price](#) | [State](#) | [Year](#) |

[Ford Fairlane Cars For Sale Classified Feed](#)

1966 Ford Fairlane


 This is a beautiful Fairlane 500 GT convertible in excellent condition. This car is straight and solid with excellent red exterior that was clearly a ... [more...](#)

\$47,000

Location:

Posted: 2007-08-17

1966 Ford Fairlane

 EVERY FORD ENTHUSIAST IN THE COUNTRY SHOULD BE ON OUR WEBSITE, VIEWING THE LARGEST THOUROUGHBRED INVESTMENT QUALITY FORD INVENTORY IN THE HISTORY OF O... [more...](#)

\$32,000

Location: , NONE

Posted: 2006-05-16

Show these listings on your website
Display the most recently added 1966 Ford Fairlane classifieds on your website.
Simply copy the code below, and paste it onto the page where you'd like the listings to show:
Ford Fairlane listings from

[What's This?](#)

Here's value g that'll fi your po
A quick & reference carry whe you go.
Lists values of luxury, exotic cars & trucks 2007. Older trucks 1946-1966. 274 pages, s
ONLY \$99 (or by subs updated ev
CLICK

Embedding :: ReM Transparency

The screenshot shows a web browser window displaying an arXiv.org page. The browser's address bar shows the URL: `http://www.openarchives.org/ore/0.2/discovery-examples/arxiv.html`. The page title is "[astro-ph/0601007] Parametrization of K-essence and Its Kinetic Term". The page content includes the title "Parametrization of K-essence and Its Kinetic Term" by Hui Li, Zong-Kuan Guo, and Yuan-Zhong Zhang. It also features a "Download:" section with links for PostScript, PDF, and Other formats, and a "References & Citations" section with links to SLAC-SPIRES HEP, NASA ADS, and CiteBase. A "Submission history" section is also present. A red box highlights a "Resource Map" section at the bottom, which includes a link back to arXiv and a URL: `http://arxiv.org/rem/astro-ph/0601007`. The browser's search bar at the bottom shows "Find:" and "Done".

Browser window: [astro-ph/0601007] Parametrization of K-essence and Its Kinetic Term
Address bar: `http://www.openarchives.org/ore/0.2/discovery-examples/arxiv.html`
Search: Google

arXiv.org > astro-ph > arXiv:astro-ph/0601007

Astrophysics

Parametrization of K-essence and Its Kinetic Term

Hui Li, Zong-Kuan Guo, Yuan-Zhong Zhang
(Submitted on 31 Dec 2005 (v1), last revised 18 Jan 2006 (this version, v2))

We construct the non-canonical kinetic term of a k-essence field directly from the effective equation of state function $w_k(z)$, which describes the properties of the dark energy. Adopting the usual parametrizations of equation of state we numerically reproduce the shape of the non-canonical kinetic term and discuss some features of the constructed form of k-essence.

Comments: 8 pages, 1 figure; accepted by Mod. Phys. Lett. A; minor changes to references
Subjects: Astrophysics (astro-ph)
Journal reference: Mod.Phys.Lett. A21 (2006) 1683-1690
DOI: [10.1142/S0217732306019475](https://doi.org/10.1142/S0217732306019475)
Cite as: [arXiv:astro-ph/0601007v2](https://arxiv.org/abs/astro-ph/0601007v2)

Submission history

From: Hui Li [[view email](#)]
[v1] Sat, 31 Dec 2005 04:01:23 GMT (20kb)
[v2] Wed, 18 Jan 2006 06:16:15 GMT (20kb)

Link back to: [arXiv](#), [form interface](#), [contact](#).

Resource Map for arXiv:astro-ph/0601007 <http://arxiv.org/rem/astro-ph/0601007>
(What's a Resource Map?)

Find: Next Previous Highlight all
Done

Embedding :: HTTP Response

```
HEAD http://www.example.net/hello.jpeg HTTP/1.1
Host: www.example.net
Connection: close
```

```
HTTP/1.1 200 OK
Date: Sat, 26 May 2007 22:43:10 GMT
Server: Apache/2.2.0
Last-Modified: Sat, 26 May 2007 19:32:04 GMT
ETag: "c3596-816-92123500"
Accept-Ranges: bytes
Content-Length: 2070
Link: <http://example.net/hw.atom>; type="application/atom+xml"; rel="resourcemap"
Content-Type: image/jpeg
Connection: close
```

Nottingham's IETF Draft establishing semantic equivalence
between HTML `<link>` and HTTP `Link`:

How Not to Do It

- Proscriptive as well as prescriptive...
 - ReMs are for machines, not humans
 - avoiding URI ambiguity

Bad :: ReMs not for Humans

```
<html>
```

```
...
```

```
<h1>Welcome to my happy page of ReMs!</h1>
```

```
<a href="http://www.foo.edu/objects/object1.atom">ReM 1</a>
```

```
<a href="http://www.foo.edu/objects/object2.atom">ReM 2</a>
```

```
<a href="http://www.foo.edu/objects/object3.atom">ReM 3</a>
```

```
...
```

```
</html>
```

Danger: You can end up confusing your users.

Yes, ReMs are 1st class resources, but normal people (present company excluded, of course) do not enjoy reading raw XML.

Bad :: URI Conflation

RFC 2295 Style Content Negotiation:

(ReM)	<code>http://www.foo.edu/objects/object1.atom</code>
(Splash Page)	<code>http://www.foo.edu/objects/object1.html</code>
(Conflated URI)	<code>http://www.foo.edu/objects/object1</code>

HTTP 303 Redirection:

(ReM)	<code>http://www.foo.edu/data/objects/object1</code>
(Splash Page)	<code>http://www.foo.edu/page/objects/object1</code>
(Conflated URI)	<code>http://www.foo.edu/resource/objects/object1</code>

`danger 1: Report 12`
`danger 2: Conflated-URI somePredicate someObject`
Is the HTML link or triple about the ReM or the Splash Page?
Depends on who is asking...

URI Conflation :: Open Issue

Allowed: Splash Page = ReM + XSLT

Why: URI-R is still returning only a ReM

From Section 5.2:

Note that these restrictions do not prevent a ReM from being used as a the basis or "ingredient" of a splash page. Servers MAY choose to include stylesheets with ReMs to make them suitable for use by human agents. Although this is an option, clients should note that there is no requirement for ReMs and splash pages to be transformable from one to another; a ReM may not have the same URIs as a splash page and vice versa.

Open Issue: ReMs in RDFa/Microformats in Splash Pages

Why Maybe Bad: URI-R is returning 2 things mixed together

Why Maybe OK: Every client gets the same 2 things from URI-R

weird but not wrong triple:

```
index.html#aggregation ore:aggregates index.html
```

don't lose the "#aggregation", or you get:

```
index.html ore:aggregates index.html
```

Discovery is a Dirty Job



- Frequently a trade-off between “cleanliness” and “utility”
- Multiple discovery methods, possibly more evolving over time
- Each method has caveats and multiple opportunities to get it wrong
- At least 2 open issues, perhaps more that we have yet to uncover