

Interoperability and Its Role In Standardization

Peter Murray

Assistant Director, Technology Services Development, LYRASIS

NISO Update, American Library Association 2013



Cape Town Book Fair at Cape Town Internation Convention Centre by [Danie van der Merwe](#) on Flickr





Masterton District Library by [Keith Davenport](#) on Flickr



The OCLC Mothership by [Michael Porter](#) on Flickr

checkout



Self checkout stations by [San José Library](#) on Flickr



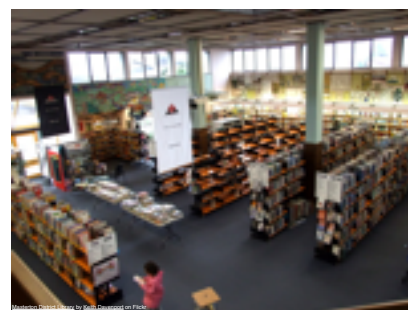
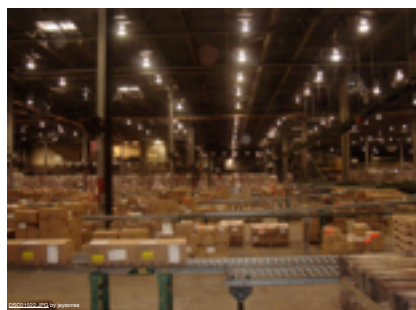
The Discharge Area by [Canton Public Library](#) on Flickr



The future of books by [Johan Larsson](#) on Flickr



eBook Readers Galore by Michael Porter on Flickr





[The ghost of a systems admin](#) by [Mike Wallis](#) on Flickr

Tuesday, July 2, 13

Standards Making It Happen

- MARC
- ONIX
- Z39.50
- EPub
- SIP/NCIP
- EDI
- ISBN
- SUSHI
- OpenURL
- DOI
- SRU/SRW
- ERM
- METS
- HTML
- RDF
- OAI-PMH
- DAISY
- PDF

Four Levels of Interoperability

“Three dimensions of organizational interoperability”, European Journal of ePractice, N° 6 · January 2009 · ISSN: 1988-625X

NISO Update, American Library Association, 30-Jun-2013

Four Levels of Interoperability

Layer	Aim	Objects	Solutions
-------	-----	---------	-----------

“Three dimensions of organizational interoperability”, European Journal of ePractice, N° 6 · January 2009 · ISSN: 1988-625X

NISO Update, American Library Association, 30-Jun-2013

Four Levels of Interoperability

Layer	Aim	Objects	Solutions
Technical Interoperability	Technically Secure Data Transfer	Signals	Protocols of Data Transfer

“Three dimensions of organizational interoperability”, European Journal of ePractice, N° 6 · January 2009 · ISSN: 1988-625X

Four Levels of Interoperability

Layer	Aim	Objects	Solutions
Technical Interoperability	Technically Secure Data Transfer	Signals	Protocols of Data Transfer
Syntactic Interoperability	Processing of Received Data	Data	Standardized Data Exchange (e.g. XML)

“Three dimensions of organizational interoperability”, European Journal of ePractice, N° 6 · January 2009 · ISSN: 1988-625X

Four Levels of Interoperability

Layer	Aim	Objects	Solutions
Technical Interoperability	Technically Secure Data Transfer	Signals	Protocols of Data Transfer
Syntactic Interoperability	Processing of Received Data	Data	Standardized Data Exchange (e.g. XML)
Semantic Interoperability	Processing and Interpretation of Data	Information	Common Directories, Data Keys, Ontologies

“Three dimensions of organizational interoperability”, European Journal of ePractice, N° 6 · January 2009 · ISSN: 1988-625X

Four Levels of Interoperability

Layer	Aim	Objects	Solutions
Technical Interoperability	Technically Secure Data Transfer	Signals	Protocols of Data Transfer
Syntactic Interoperability	Processing of Received Data	Data	Standardized Data Exchange (e.g. XML)
Semantic Interoperability	Processing and Interpretation of Data	Information	Common Directories, Data Keys, Ontologies
Organizational Interoperability	Automatic Linkage of Processes	Process (Workflow)	Standardized Process Elements

“Three dimensions of organizational interoperability”, European Journal of ePractice, N° 6 · January 2009 · ISSN: 1988-625X

Today...

- **ResourceSync**
Peter Murray, LYRASIS
- **NCIP**
Mike Dicus, Ex Libris
- **SIP**
Ted Koppel, Auto-Graphics
- **IOTA**
Rafal Kasprowski, Rice University
- **Other NISO Initiatives**
Nettie Lagace, NISO



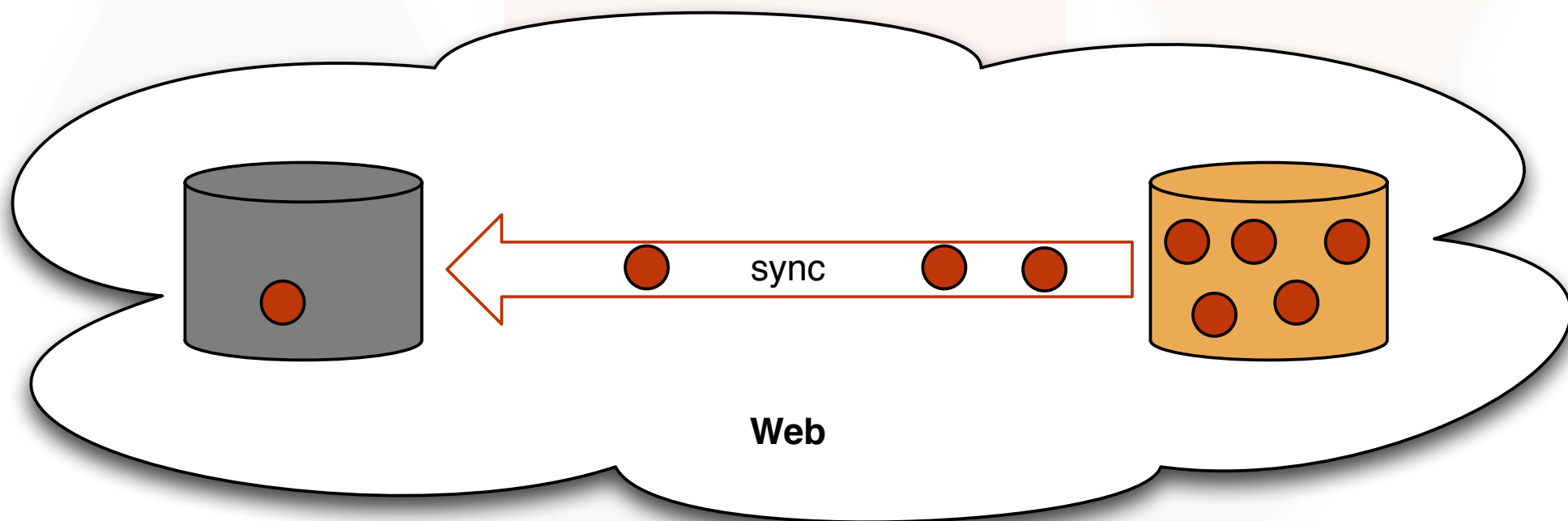
<http://www.openarchives.org/rs/>

ResourceSync: Leveraging Sitemaps for Resource Synchronization

Slides adapted from talk by
Bernhard Haslhofer, University of Vienna
WWW 2013, Rio de Janeiro, May 17th

What?

- A framework for synchronizing **Web resources** from a **Source** to a **Destination**



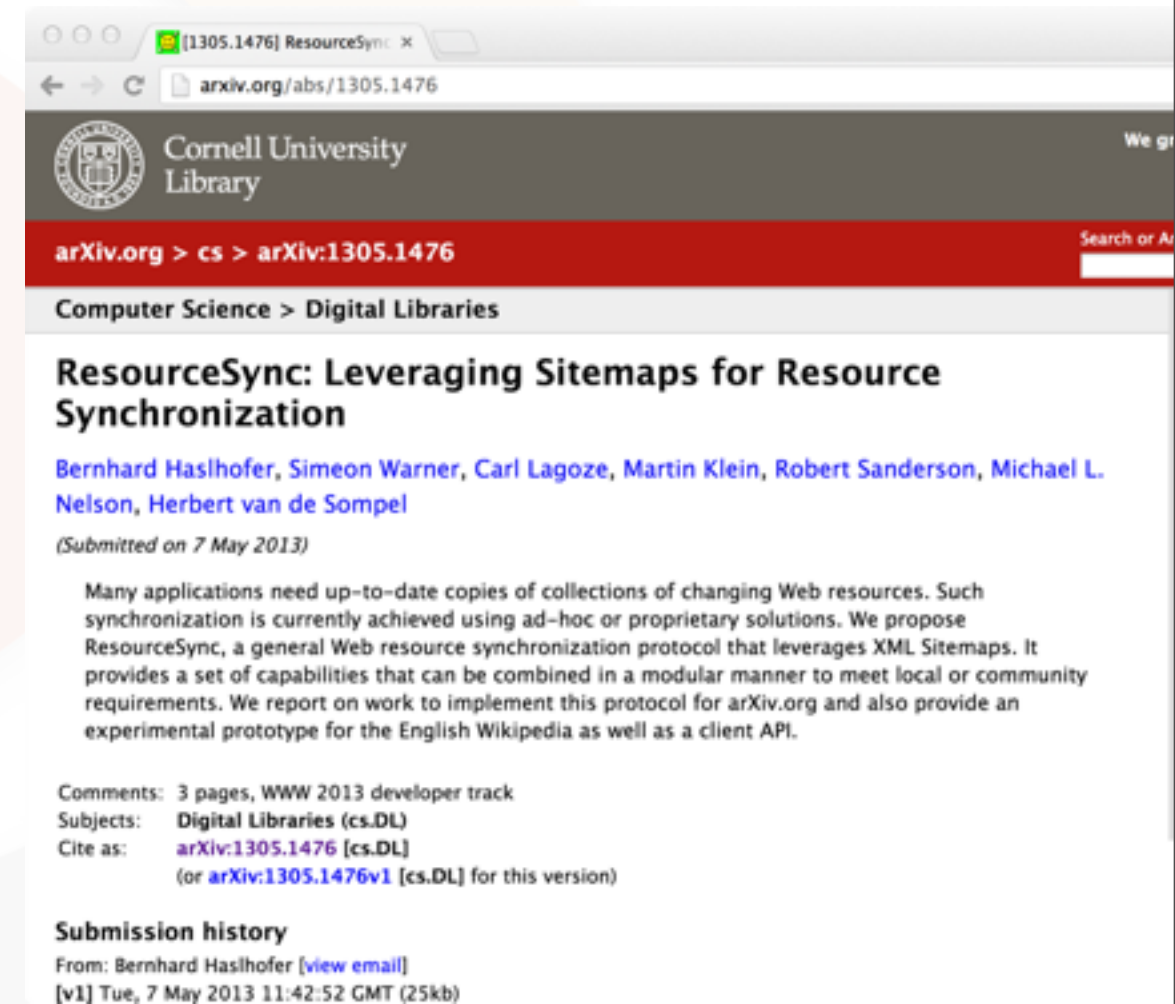
Why?

- **rsync**: filesystem sync, but not Web
- **OAI-PMH**: metadata, but not resources
- **Web-DAV**: extends HTTP, requires server installation at source
- ...

... because lots of projects and services are doing synchronization but rely on ad-hoc solutions!

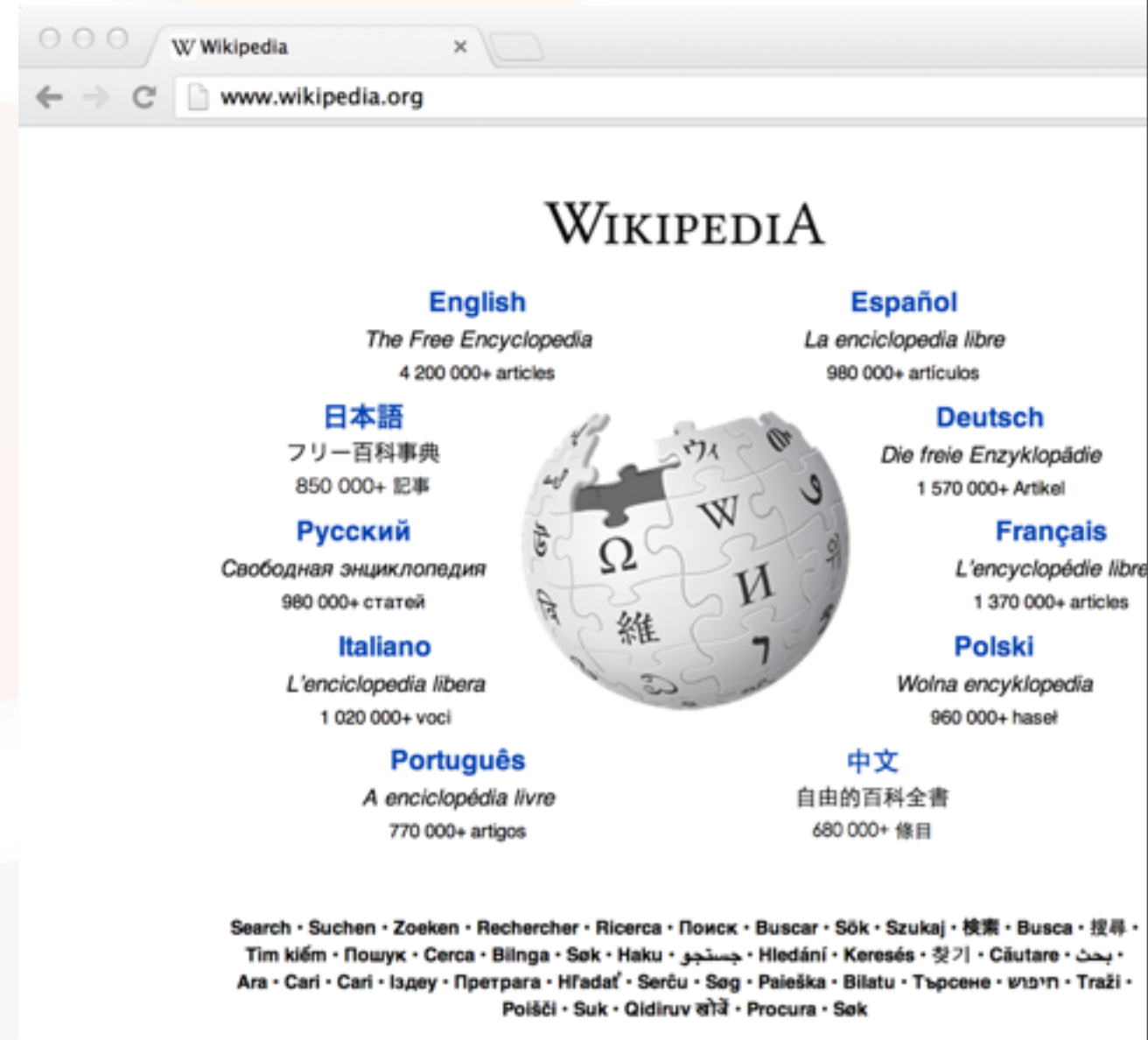
arXiv.org mirroring

- 2.4M resources (PDF, metadata, Latex src)
- ~800/day created or updated
- uses homebrew mirroring since 1994 (!)
- look for more general solution to support independent destinations



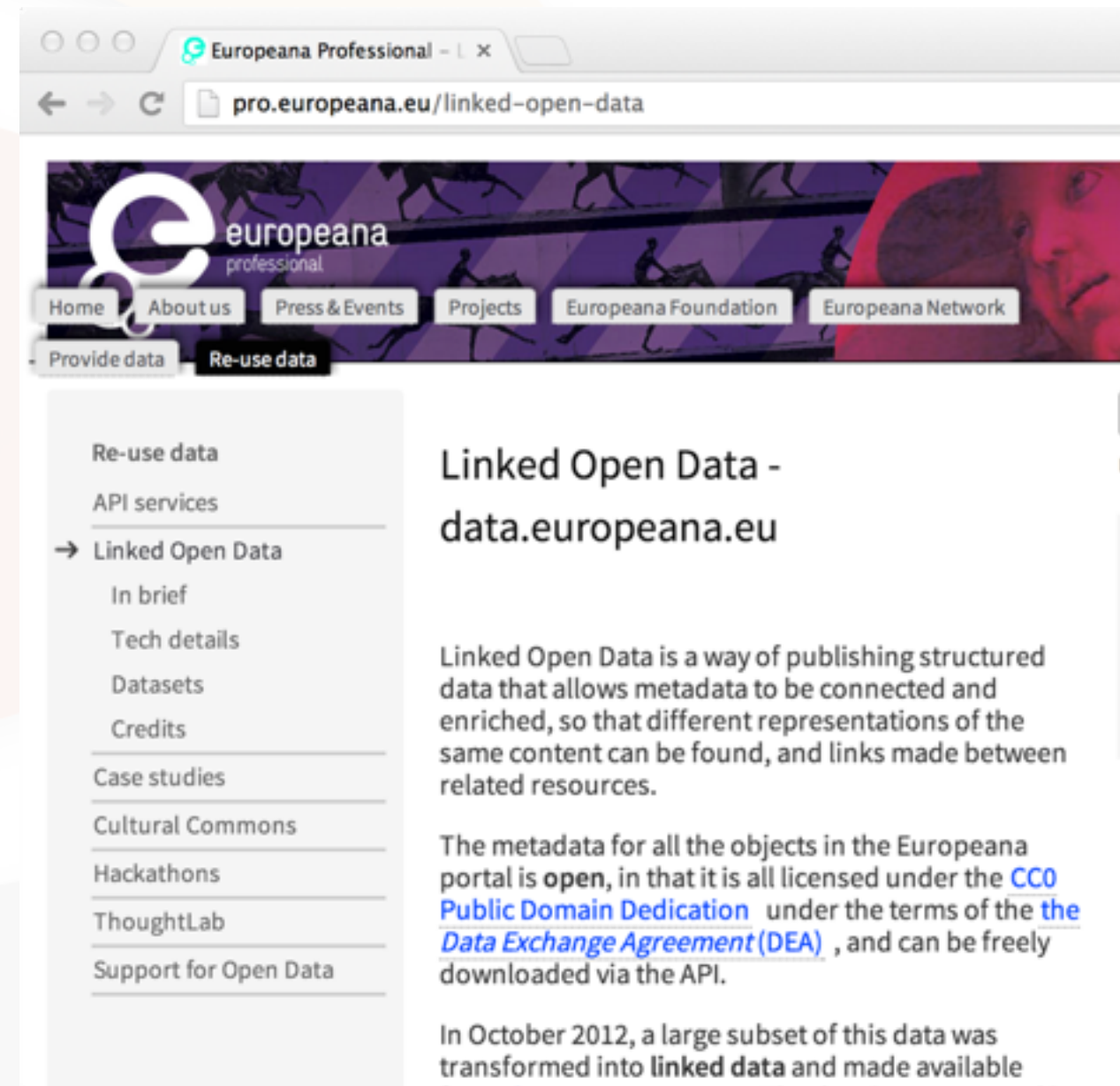
Wikipedia

- 1.4 updates / sec
- many dependent services reusing Wikipedia content (e.g., DBPedia, Freebase, etc.)
- harvest articles via OAI-PMH, retrieve changes via IRC, download dumps



data.europeana.eu

- aggregates metadata from >200 data providers in Europe
- 10 largest providers contribute 80%
- >190 providers contribute 20%



Design Guidelines

- Sync **small websites** / repositories (few resources) but also **large data collections** (millions of resources)
- Support **low change frequency** (weeks / months) to **high change frequency** (seconds) sources
- **Low adoption barrier!**

Builds on 'Sitemaps' Specs

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
        xmlns:rs="http://www.sitemaps.org/rs"
        capability="urlcelist"
        from="2013-01-02T09:00:00Z">
  <url>
    <loc>http://example.com/res1</loc>
    <lastmod>2013-01-02T13:00:00Z</lastmod>
    <rs:md hash="md5:1584abdf8ebdc98c0c6a7402">
  </url>
  <url>
    <loc>http://example.com/res2</loc>
    <lastmod>2013-01-02T14:00:00Z</lastmod>
    <rs:md hash="md5:1e0d5eb0e16ba40c99b14c02">
  </url>
</urlset>
```

Builds on 'Sitemaps' Specs

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
        xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:md capability="resourcelist"
        from="2013-01-03T09:00:00Z"/>
  <url>
    <loc>http://example.com/res1</loc>
    <lastmod>2013-01-02T13:00:00Z</lastmod>
    <rs:md hash="md5:1584abdf8ebdc9802ac0c6a7402c03b6"/>
  </url>
  <url>
    <loc>http://example.com/res2</loc>
    <lastmod>2013-01-02T14:00:00Z</lastmod>
    <rs:md hash="md5:1e0d5cb8ef6ba40c99b14c0237be735e"/>
  </url>
</urlset>
```

Builds on 'Sitemaps' Specs

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.sitemaps.org/rs/sitemap/0.9"
  capability="image:image"
  from="2013-01-01T11:00:00Z"
  until="2013-01-03T11:00:00Z"/>

  <url>
    <loc>http://example.com/res2.pdf</loc>
    <lastmod>2013-01-02T13:00:00Z</lastmod>
    <rs:change>updated</rs:change>
  </url>
  <url>
    <loc>http://example.com/res3.tiff</loc>
    <lastmod>2013-01-02T18:00:00Z</lastmod>
    <rs:change>deleted</rs:change>
  </url>
</urlset>
```

Builds on 'Sitemaps' Specs

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
        xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:md capability="changelist"
        from="2013-01-01T11:00:00Z"
        until="2013-01-03T11:00:00Z" />
  <url>
    <loc>http://example.com/res2.pdf</loc>
    <lastmod>2013-01-02T13:00:00Z</lastmod>
    <rs:md change="updated" />
  </url>
  <url>
    <loc>http://example.com/res3.tiff</loc>
    <lastmod>2013-01-02T18:00:00Z</lastmod>
    <rs:md change="deleted" />
  </url>
</urlset>
```

	Baseline Synchronization	Incremental Synchronization	Audit
• URI • Metadata - fixity - links	Resource List	Change List	Resource List fixity Change List fixity
• URI • Bitstream • Metadata - fixity - links	Resource Dump	Change Dump	Resource Dump fixity Change Dump fixity

Status

- Beta spec (v. 0.9) for public comment
<http://www.openarchives.org/rs/toc>
- Tool development started
- Separate documents for archiving and push deployments

Next Steps

- Continue tool development & deployment
- Collect
 - public comments on resourcesync@googlegroups.com
 - implementation issues on <https://github.com/resync/resync/issues>
- Version 1.0 in fall 2013 (NISO standard)

Thanks!

@DataG

<http://slideshare.net/DataGazetteer>

<http://openarchives.org/rs>

resourcesync@googlegroups.com