Interoperability and Its Role In Standardization

Peter Murray
Assistant Director, Technology Services Development, LYRASIS

NISO Update, American Library Association 2013
Self checkout stations by San José Library on Flickr
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
## Four Levels of Interoperability

<table>
<thead>
<tr>
<th>Layer</th>
<th>Aim</th>
<th>Objects</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical</td>
<td>Secure Data Transfer</td>
<td>Signals</td>
<td>Protocols of Data Transfer</td>
</tr>
<tr>
<td>Syntactic</td>
<td>Processing of Received Data</td>
<td>Standardized Data Exchange (e.g. XML)</td>
<td>Processing and Interpretation of Data</td>
</tr>
<tr>
<td>Organizational</td>
<td>Automatic Linkage of Processes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Four Levels of Interoperability

<table>
<thead>
<tr>
<th>Layer</th>
<th>Aim</th>
<th>Objects</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Interoperability</td>
<td>Technically Secure Data Transfer</td>
<td>Signals</td>
<td>Protocols of Data Transfer</td>
</tr>
</tbody>
</table>

## Four Levels of Interoperability

<table>
<thead>
<tr>
<th>Layer</th>
<th>Aim</th>
<th>Objects</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical Interoperability</strong></td>
<td>Technically Secure Data Transfer</td>
<td>Signals</td>
<td>Protocols of Data Transfer</td>
</tr>
<tr>
<td><strong>Syntactic Interoperability</strong></td>
<td>Processing of Received Data</td>
<td>Data</td>
<td>Standardized Data Exchange (e.g. XML)</td>
</tr>
</tbody>
</table>
## Four Levels of Interoperability

<table>
<thead>
<tr>
<th>Layer</th>
<th>Aim</th>
<th>Objects</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical Interoperability</strong></td>
<td>Technically Secure Data Transfer</td>
<td>Signals</td>
<td>Protocols of Data Transfer</td>
</tr>
<tr>
<td><strong>Syntactic Interoperability</strong></td>
<td>Processing of Received Data</td>
<td>Data</td>
<td>Standardized Data Exchange (e.g. XML)</td>
</tr>
<tr>
<td><strong>Semantic Interoperability</strong></td>
<td>Processing and Interpretation of Data</td>
<td>Information</td>
<td>Common Directories, Data Keys, Ontologies</td>
</tr>
</tbody>
</table>

## Four Levels of Interoperability

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

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
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.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="md5:1584abdf8ebdc988c0c6a7402c0f500b"/>
    </url>
    <url>
        <loc>http://example.com/res2</loc>
        <lastmod>2013-01-02T14:00:00Z</lastmod>
        <rs:md hash="md5:1e0d5ebd8e16ba40c99b14c032a0a0b"/>
    </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.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>
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/"
    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>
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