Distributed Authority

Hugh Glaser
Seme4 Ltd.

With thanks to many, including Ian Millard
• What Happens When Authorities Collide?
• What is the Granularity of an Authority?
• When we have PIDs, what do we do next?
• How do we Manage the Infrastructure?

• http://eprints.ecs.soton.ac.uk/18452/
Some Assumptions

• Authority is established, not bestowed
• IDs are URIs and can be resolved
  – Preferably to RDF, but not necessarily
• But they must resolve
  – otherwise you can’t even guess what they mean
• Not just people
  – the infrastructure should work for publications, institutions, lectures, experiments...
• We want to solve problems, not create them
• I don’t have long...
Barcode on the forehead is a Chimera
Each Metadata owner publishes its own identifiers

• **This is absolutely right**
  – Their workflow depends on them
  – Their concept of the entity will be unique to them
  – They own it
  – If they used someone else’s, how would they deal with a change?
    • Split of entity
    • Join of entity
    • Correction of entity
Institutional “Use cases”

• What other institutions do my workers work with?
• What do my departments know about Jane?
• What projects (RCUK, JISC, EU) do we have in photovoltaics?
  – Why would a University keep project metadata?
    • Surely they can just “use” the JISC metadata
• Give me a snapshot of Jane’s activity
• How does the library know who is using which texts?
Identity Knowledge

• Is different from other knowledge

• Goodness about separation
  – License
  – Provenance
  – Trust
  – Duplication (maintenance)
  – Internal process

• Mistake to combine discovery with service
  – Simple services that do one set of things well
  – Do not duplicate/copy data, e.g. names
    • Creates many problems: coherence, maintenance...

• Identity is context-dependent
  – Use context more than publication context
Some requirements

• **Scalability**
  - Distributed
  - Metadata publishers should be able to give some authority of their equivalence

• **These may be context dependent, e.g.**
  - Equivalence as a worker here (Department, JISC?)
  - Equivalence as a professional (previous inst.)
  - Equivalence as a person (dbpedia)

• **Other should be able to make different statements**

• **Essentially anarchy, as the rest of the Web**
Co-Reference Service (CRS)

- How we do it
- RESTful service
- URI<sub>i</sub> -> { URI<sub>1</sub>, ..., URI<sub>i</sub>, ..., URI<sub>n</sub> }
- Any metadata owner can publish co-reference to other sources
  - Multiple “opinions” if required
- Others can provide such services
- Middleware or simpler can aggregate
  - From recognised/approved CRSes
  - “Let the buyer beware”?
  - “Let the buyer be empowered”
For Example

- JISC: Metadata, CRS
- Southampton: Metadata, ePrints, CRS: Public, CRS: Internal
- HEFCE: Metadata, CRS: People at institutions, CRS: People
- ISI: CRS: Lite (free), CRS: Subscription

Seme4: Linked Data Solutions
Some Other Issues

- Deprecation
- CRS Discovery
  - Recommended CRSes from metadata from resolved URI
- Aggregating Services
- Robustness to failure
- Actually CRS is not confined to Linked Data
Examples

• Running out of time
• For the Sceptical
• So this is what can be done
  – http://rkbexplorer.com/
  – http://dotac.info/
  – http://sameas.org/
  – http://jisc.rkbexplorer.com/id/org-77
A Web-Based Resource Model for eScience: Object Reuse & Exchange

Related publications:
- Pathways: augmenting L....
- An information network...
- Core Services in the A....
- E-points and Journal A....
- An Architecture for th....
- Web-Based Resource M....
- Object Re-Use & Exchan...
- Access Interfaces for ...

Details:
Title: A Web-Based Resource Model for eScience: Object Reuse & Exchange
Author: Herbert Van de Sompel, Simeon Warner, Peter Johnstone, Robert Sanderson
Web Address: http://www.eks.org/abs/0811.0573
Journal: Volume: abs/0811.0573

Publications:
- Object Re-Use & Exchange: A Resource-Centric Approach
- An Architecture for the Large-Scale Modeling of Scholarly Artifacts and their Usage
- Access Interfaces for Open Archival Information Systems based on the OAI-PMH and the OpencITI Framework for Context-Sensitive Services

Research Areas:
No results found
And how they are linked
And Why are they Linked?

www.rkbexplorer.com

Carl Lagoze is connected to Dean Krafft

They are linked by 9 relations:

- Core services in the architecture of the national science digital library (NSDL)
- An information network overlay architecture for the NSDL
- Metadata aggregation and "automated digital libraries": A retrospective on the NSDL experience

Affiliations

They are both affiliated to Cornell University.

Projects

They are both members of 2 projects:

- NSDL Technical Network Services: A Cyberinfrastructure Platform for STEM Education
- Collaborative Project: Core Integration - Leading NSDL toward Long-Term Success
Complete Co-Reference Information

This service computes the equivalence class within the known URIs for a specified URL by consulting all relevant CRS knowledge bases.

Equivalent URIs...
1. (Canon) http://acm.kbexplorer.com/id/person-407157
4. http://dblp.rkbexplorer.com/id/people-tec5a6002982222d0374695a52140590d423eb148125a6e565738bc1a143b

The following diagram shows the interconnectivity between the CRS knowledge bases which maintain the context-dependent representation of co-reference for each of the RKBExplorer domains.

Seungwoo Lee

Showing information queried from all repositories ...
Using the CRS Service

Canon
http://jisc.rkbexplorer.com/id/org-77

Duplicate URLs
http://dotac.rkbexplorer.com/id/org-cumbria-ac-uk
http://jisc.rkbexplorer.com/id/org-348
http://jisc.rkbexplorer.com/id/org-72
http://jisc.rkbexplorer.com/id/org-77

Bundle instantiated at
2009-11-28 14:06:29

Alternative representations
RDF/XML
sameAs list in N3
image/png
Solving an Existing Problem

University of Cumbria

Showing combined information from the following coreferenced URIs ...

http://jisc.rkbexplorer.com/id/org-348 [View this URI only]
http://jisc.rkbexplorer.com/id/org-72 [View this URI only]
http://jisc.rkbexplorer.com/id/org-77 [View this URI only]

The CRS managing coreference data for jisc.rkbexplorer.com knows of 1 additional equivalent URIs in other repositories.
You can also view the global equivalence closure across all repositories.

Results in repository jisc.rkbexplorer.com...

<table>
<thead>
<tr>
<th>Subject</th>
<th>Property</th>
<th>Object/Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumbria Institute of the Arts</td>
<td>jisc:name</td>
<td>Cumbria Institute of the Arts</td>
<td>jisc-institutions.rdf</td>
</tr>
<tr>
<td>St Martin's College</td>
<td>jisc:name</td>
<td>St Martin's College</td>
<td>jisc-institutions.rdf</td>
</tr>
<tr>
<td>University of Cumbria</td>
<td>jisc:name</td>
<td>University of Cumbria</td>
<td>jisc-institutions.rdf</td>
</tr>
<tr>
<td>University of Cumbria</td>
<td>rdf:type</td>
<td>foaf:Organization</td>
<td>jisc-institutions.rdf</td>
</tr>
<tr>
<td>Cumbria Institute of the Arts</td>
<td>rdf:type</td>
<td>foaf:Organization</td>
<td>jisc-institutions.rdf</td>
</tr>
<tr>
<td>St Martin's College</td>
<td>rdf:type</td>
<td>foaf:Organization</td>
<td>jisc-institutions.rdf</td>
</tr>
</tbody>
</table>
Finding Co-reference

<sameAs>
interlinking the Web of Data

The Web of Data has many equivalent URIs. This service helps you to find co-references between different data sets. Enter a known URL or use Sindice to search first.

Search results from Sindice, with co-references applied...

Les Carr

3. http://southampton.rkbexplorer.com/fd/person-a635f51b18c81b1f1df7dab..84425e0250

Show 118 more

Les Carr

1. http://hblp.l3s.de/d2rhresource/authors/Les_Carr

Currently serving 23280924 URLs in 9454882 bundles!
Mark Borkum did this
Concluding Remarks

- CRS
- A service to mediate between authorities
- Simple, mature service (5 years+)
- Well understood, already deployed
- Can be used now, without major coordination
- Open

We have [http://oai.rkbexplorer.com/](http://oai.rkbexplorer.com/)
What Might JISC Do?

- Probably clear what I think
  - Facilitate CRS deployment
    - Software
      - Package CRS software
      - (They funded a co-reference editor as part of dotAC)
    - Training
    - Outreach
  - Take co-reference discovery initiatives
    - Publish co-reference data in a CRS
  - Make their data available as Linked Data
    - Examples are good
- Join the data.gov.uk world
- Encourage/help other to do the same
URIs

- http://sameas.org/
- http://jisc.rkbexplorer.com/id/org-77
- http://rkbexplorer.com/
- http://dotac.info/
- http://oai.rkbexplorer.com/
- http://eprints.ecs.soton.ac.uk/18452/