UK ORE Information Day
Open Repositories 08, 04/04/08

Functional Object Re-use and Exchange: Supporting Information Topology Experiments

http://foresite.cheshire3.org/

Rob Sanderson (azaroth@liverpool.ac.uk)
Richard Jones (richard.d.jones@hp.com)
Clare Llewellyn (clare.llewellyn@jstor.org)

UNIVERSITY OF LIVERPOOL

Functional Object Re-use and Exchange: Supporting Information Topology Experiments

Overview:

- Introduction
- Scholarly Communication
- JSTOR
- Project Overview
- ORE Journal Descriptions
- DSpace
- SWORD
- Future Work?

Introduction

Experiments to provide feedback on ORE specifications

Two experiments in the UK with JISC funding

Lots of possibilities, but scholarly communication most appropriate for JISC.

Two phases:

Describe journal/issue/article hierarchy in ORE

Import descriptions into DSpace

Grand Vision:

Bootstrap ORE-based scholarly communication processes!

Scholarly Communication

ORE has many advantages for Scholarly Communication:

- Easy interoperability via ATOM
  - Plays well in the current environment
- Built on Web Architecture
- Solid and well specified abstract model
- Network of relationships easily described in RDF
- Proxies allow for new and useful constructions
- Aggregation as representation of any collection of web resources allows for seamless interaction between different communities and types of resource.
- Easy to build into open publishing / institutional repositories

Scholarly Communication

Requirements:

- Large collection of scholarly communication!
- Preferably working with data providers, rather than using openly harvestable data (eg arxiv, citeseer, pubmed)
- Collection described in such a way as to allow transformation into ORE Resource Maps

- Ingestion protocol to upload to DSpace
- DSpace to understand ORE Resource Maps
- DSpace to allow linking back to original source, rather than storing data locally

JSTOR

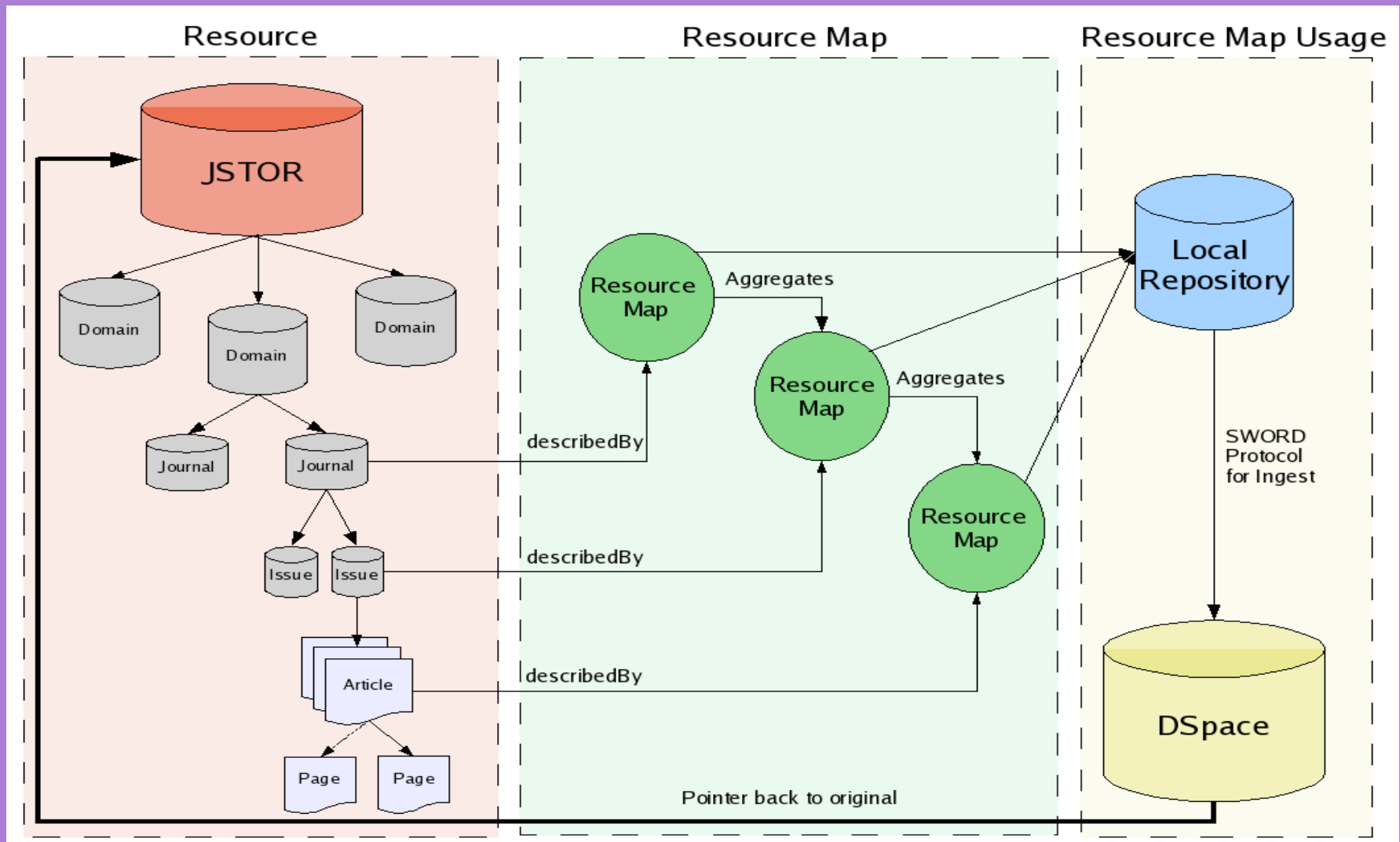
Enter JSTOR:

- Very Large collection of journals
  - 1000+ journals, 185,000+ issues, 1.8M+ articles
- Described in XML, down to OCR of article text
  - 200 Gigabytes of compressed XML
- Data not otherwise available, due to publishers' restrictions

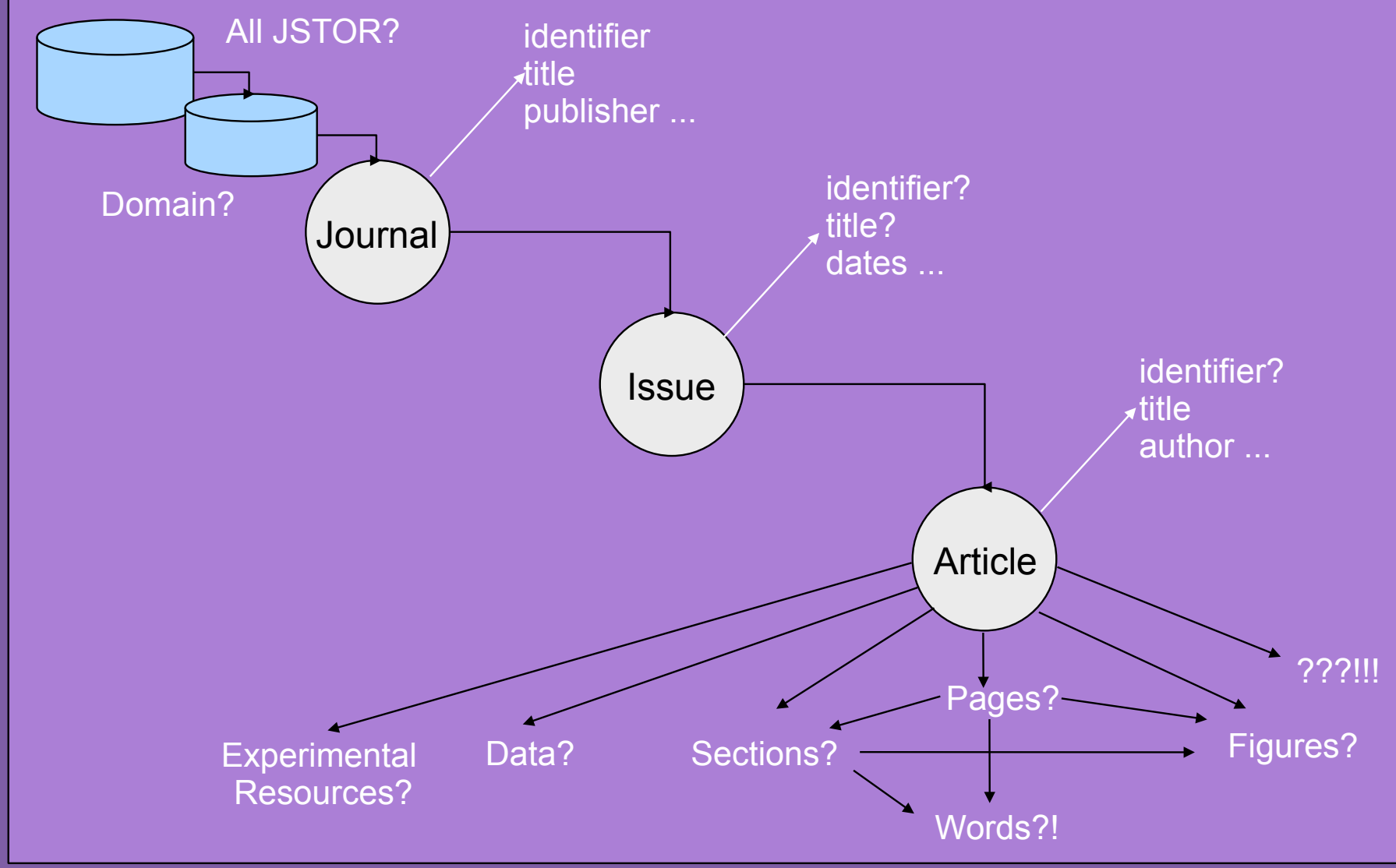
Advantages for JSTOR:

- Instant entree into Semantic Web game
- Harvestable ReMs without giving away publishers' data
- Google indexing ReMs will drive traffic to site
- Graph analysis/mining enables new functionality
- Graph Visualisation options
- Integration with scholarly communication software

Project Overview



Journal Descriptions in ORE



Obligatory Angle Brackets

```
<journal-meta>
  <journal-id journal-id-type="doi">10.2307/j100802</journal-id>
  <journal-title>Review of Financial Studies</journal-title>
  <issn pub-type="ppub">08939454</issn>
  <publisher>Oxford University Press</publisher>
 </journal-meta>
 <issue-meta>
  <copyright>Copyright 1991 The Society for Financial Studies</copyright>
  <pub-date pub-type="ppub">
     <day>1</day> <month>1</month> <year>1991</year>
  </pub-date>
  <voliss> <volume>4</volume> <issue>4</issue> </voliss>
  <issue-title/>
 </issue-meta>
 <articles>
  <article type="research-article">
   <article-id pub-id-type="doi">10.2307/2962151</article-id>
   <article-title>Asymmetric Predictability of Conditional Variances</article-title>
   <contrib contrib-type="author"><name>
     <given-names>Jennifer</given-names>  <surname>Conrad</surname>
   </name></contrib>
  ...
```

More Obligatory Angle Brackets

```
<feed>
  <id>http://foresite.cheshire3.org/jstor/j100802/ore/</id>
  <title>Review of Financial Studies</title>
  <author><name>Oxford University Press</name></author>
  <link href="info:doi/10.2307/j100802" rel="related"/>
  <link href="urn:issn:08939454" rel="related">
  <category term="http://www.openarchives.org/ore/terms/Aggregation"/>
  <link rel="self" href="http://foresite.cheshire3.org/jstor/j100802/ore/atom.xml"/>
  <updated>2008-04-02T16:00:00</updated>

  <entry>
    <id>http://foresite.cheshire3.org/jstor/j100802/ore/proxy/i352820</id>
    <title/>
    <updated>2008-04-02T16:00:00</updated>
    <link rel="alternate" href="http://foresite.cheshire3.org/jstor/j100802/i352820/ore/"/>
    <rdf:Description about="http://foresite.cheshire3.org/jstor/j100802/i352820/ore/">
      <rdf:type>http://www.openarchives.org/ore/terms/Aggregation</rdf:type>
    </rdf:Description>
  </entry>
  ...
</feed>
```

DSpace

No need to discuss DSpace in general!

DSpace Development Tasks from Foresite:

- Allow Resource Maps to be submitted via SWORD
- Storage of Resource Maps
- Re-Identifier-ing (?!) of Resource Maps once ingested
- Dereferencing resources in Resource Maps
- Allowing pointers to Resources instead of dereferencing
- Returning ORE/SWORD responses

SWORD

SWORD: Simple Web-service Offering Repository Deposit

JISC funded between 1 March and 31 October 2007
Profile of Atom Publishing Protocol

Simple Case:

Repository publishes self-describing service document
Client POSTs data (with HTTP headers) to Repository
Repository responds with an <atom:entry> document

Less Simple Case:

Client POSTs data on behalf of user to Repository
Repository authenticates and responds with
(more complex) <atom:entry> document

SWORD

Primary Scenario:

ORE as specification of compound object to be deposited:

Client POSTs ORE description to server

Server accepts, dereferences Aggregated Resources, creates object and returns response

In this case, instead of dereferencing, we point back to original JSTOR URI for the resource

Other Possible Scenarios:

POST zip of resources plus ORE description to server

POST to Proxy URI (atom:entry/atom:id) to create, update or delete an Aggregated Resource

Return SWORD+ORE (SWORED?) entries

Future Work

Well...

All of the previous slides... Project only just started!

That said...

Investigate further SWORD based scenarios
Investigate inter-repository transferd via ORE/SWORD
Repeat for open data sets (arxiv, citeseer, medline, etc)
Investigate external citation linking
Push to JSTOR's sandbox for the world to play with
Investigate graph analysis of resources

Thank You

Thank You :)

Questions?