Augmenting Institutional Repository Metadata by Harvesting Community Annotations

Annotation Systems

- **Purpose:**
  - “Value-adding”/interpretation of digital objects/scientific data
  - Secondary metadata, semantic tagging
  - Facilitate group discourse and harness collective knowledge
  - Share assertions
    - amongst trusted colleagues distributed across institutions
    - across distributed annotation servers

- **Usage:**
  - Education
  - eResearch/eScience Collaboratories
  - Peer-review

- **Applications**
  - RQF
  - Peer review systems
  - PictureAustralia
  - Scientific Data Quality Control

Jane Hunter,
University of Qld
Existing Clients

- Amaya (Client) + Annotea (Server)
- Annozilla
- Discipline-specific & media-specific systems
  - Mindswap, AstroDAS, BioDAS
- Proprietary systems
  - Word (comments)
  - PDF (comments)

Limitations of Current Systems

- Don’t support search and retrieval across distributed servers
- Lack of security/privacy mechanisms
- Lack of responsiveness
  - Slow, no notification method
- Limited search capabilities
- Limited media types supported
- Coarse granularity
- Unstructured annotations
  - Flat text only
Objectives

1. An Annotation system that supports:
   - New types of scientific data - high quality video, animations, simulation sharing, models (3D)
   - Annotation of relationships between objects
   - Compound OAI/ORE objects
   - Distributed teams of users
   - Access grid nodes, skype
   - Authentication and access control to sessions and annotations

2. The ability to harvest the annotations to augment repository metadata and search services

Co-Annotea

- Developed 2 Annotea clients
  - Browser plug-in
  - Vannotea – collaborative multimedia
- Single-sign on to distributed annotation servers
- Secure access and controlled sharing
- Source is authenticated -> trusted annotations
- Can annotate text, HTML, images, video, audio, 3D
  - Fine-grained regions/ segments
- Structured annotations
  - ranking, comments, related resources, keywords
- Annotation content - textual, URLs or files
- Annotate relationships between multiple objects -> OAI-ORE
- Search, browse and retrieval
  - across metadata (creator, date, language) and content
- Staggered, threaded – discussions
- RSS feeds – notify of changes, responses
- Attach Creative Commons licenses to annotations
Secure, Authenticated Annotations

- 2 Levels of Protection:
  - Authentication to annotation server (Shibboleth)
  - Access constraints over annotations (XACML)
    - create, list, read, edit, delete
- Combine 3 Core Technologies:
  - Annotea
  - Shibboleth
  - XACML
Annotea

- RDF-based
  - Open
  - Extensible
  - Semantic Web

- Existing Servers:
  - Zope
  - PerlLib

Annotea Extensions

- Multimedia
- SVG drawings
- XACML policies
- Any other domain specific information

[Ref: http://www.w3.org/2001/Annotea/Papers/KCAP01]
eScience Applications

- Ethnographic media analysis
- CyberSTEM – Networked Telemicroscopy
- Protein crystallography – 3D structures
- Telemedicine – Online discussions and content sharing between clinicians

Vannotea – Collaborative Annotation and Discussion of Medical Images/Videos
A normal liver stains strongly for Sput-L and easily be mistaken for a liver with malignant hepatocellular carcinoma (HCC).
System Architecture

Web Search Interface

Harvested Metadata Store

Augmented Metadata

Periodically Harvested Metadata

OAI-PMH

OAI Repository

Community generating annotations

Authenticated Annotation Service

Shibboleth

Web Search Interface

Institutional Repositories/Metadata
Institutional metadata

Community Annotations

Annotea OAI-PHM Implementation

<table>
<thead>
<tr>
<th>Java Servlet Container (Tomcat, etc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotea service configuration descriptors</td>
</tr>
<tr>
<td>Annotea servlets and filters</td>
</tr>
<tr>
<td>Annotea (core)</td>
</tr>
<tr>
<td>Annotea Record Factory (RDF)</td>
</tr>
<tr>
<td>Annotea Datastore Driver</td>
</tr>
<tr>
<td>Jena Driver</td>
</tr>
<tr>
<td>Jena over MySQL</td>
</tr>
</tbody>
</table>
Annotation->DC Mapping

<table>
<thead>
<tr>
<th>Annotea RDF Element</th>
<th>DC Element (unqualified)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creator</td>
<td>DC:creator</td>
</tr>
<tr>
<td>Date.Created</td>
<td>DC:date</td>
</tr>
<tr>
<td>Body</td>
<td>DC:description</td>
</tr>
<tr>
<td>Annotation URI</td>
<td>DC:identifier</td>
</tr>
<tr>
<td>Target</td>
<td>DC:subject</td>
</tr>
<tr>
<td>Context</td>
<td>DC:coverage</td>
</tr>
<tr>
<td>Language</td>
<td>DC:language</td>
</tr>
<tr>
<td>Policy/License</td>
<td>DC:rights</td>
</tr>
<tr>
<td>Annotation Server URI</td>
<td>DC:source</td>
</tr>
<tr>
<td>RDF type</td>
<td>DC:type</td>
</tr>
</tbody>
</table>

("annotation" subclasses = comment, question, rating, assessment)

Annotea OAI-PHM Implementation

HTTP request: http://<host>/OAI2?verb=...

Verbs:
- Identify
- List
- Identifiers
- List Sets
- Get Record
- List Records

Client: Browser or process

OAI-PMH server

<XML response>
Enhanced Search and Retrieval

• Community knowledge -> richer metadata
• Quality assessment and peer review
• FOAF/Social networks
  – use to rank results
  – Metadata/assessments of respected colleagues ranked higher in result sets
Norman F. Nelson
Collection
Record of Visit to Mission Stations 1936
Annotate photo:
UQPLST_001_0355.jpg

Describe the photo:
e.g. Fixing cornflakes, Samuel and Alfred

Describe the people (if applicable) in the photograph:
e.g. Samuel (left) and Alfred (right)

PictureAustralia harvesting

Full web harvest once per week
OAI harvest every night

Small sites
Libraries
Archives

Universities

Museums
Galleries

Large sites
SLNSW
SLVIC

War Memorial

NLA

Image – courtesy of National Library of Australia
Acknowledgements

- Ron Chernich, UQ DART Project Manager
- Ronnie Schroeter, Vannotea
- Imran Khan
  - Annotea sidebar
  - Shibboleth,XACML interface
Reference

http://www.itee.uq.edu.au/~eresearch/