Metadata Tools for eResearch

Ron Chernich
Principal Research Fellow
University of Queensland, Australia
What is eResearch?

‘e-Research’ encapsulates research activities that use a spectrum of advanced ICT capabilities and embraces new research methodologies emerging from increasing access to:

- Broadband communications networks, research instruments and facilities, sensor networks and data repositories;
- Software and infrastructure services that enable secure connectivity and interoperability;
- Application tools that encompass discipline-specific tools and interaction tools.

Source: DEST website
## Examples at UQ eResearch Lab

<table>
<thead>
<tr>
<th>Project</th>
<th>Description</th>
<th>Metadata Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aus-e-Lit</td>
<td>Collaborative integration and annotation services for Australian literature communities</td>
<td>DC, RDF</td>
</tr>
<tr>
<td>DIAS-B</td>
<td>Annotation services for the biodiversity community</td>
<td>RDF, OAI-PMH</td>
</tr>
<tr>
<td>Health-e-Waterways</td>
<td>Data integration service for smarter collaborative water management</td>
<td>DC</td>
</tr>
<tr>
<td>ARCHER</td>
<td>Australian Research Enabling Environment framework</td>
<td>&lt;various&gt;</td>
</tr>
<tr>
<td>Virgil</td>
<td>Virtual Meeting archive and retrieval</td>
<td>DC</td>
</tr>
<tr>
<td>BioMANTA</td>
<td>Modelling and Analysis of Biological Network Activity</td>
<td>RDF, OWL</td>
</tr>
<tr>
<td>EthnoER</td>
<td>Ethnographic e-research</td>
<td>DC</td>
</tr>
<tr>
<td>Pronto</td>
<td>Ontology Registry and visualization</td>
<td>OWL</td>
</tr>
<tr>
<td>Vannotea</td>
<td>Collaborative indexing, annotation and discussion for audiovisual content over broadband networks</td>
<td>RDF</td>
</tr>
<tr>
<td>HarvANA</td>
<td>Harvesting and Aggregating Networked Annotations</td>
<td>DC, OAI-PMH, RDF</td>
</tr>
<tr>
<td>MDSR</td>
<td>Metadata Schema Registry</td>
<td>&lt;various&gt;</td>
</tr>
</tbody>
</table>
Which means...

Metadata!
Common Schemas and Protocols

- Dublin Core (DC) and the Children of Dublin Core
- Resource Discovery Framework (RDF)
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
- Annotea (RDF based annotation server protocol)
- Bespoke and proprietary schemas (e.g., TARDIS and Rigaku)
- Etc, etc (all the way down… 😊)
Common Technologies

- LAMP (Linux + Apache + MySQL + P* scripting)
- PHP, Perl, Python, shell scripting
- Java (for server side portability) + Tomcat
- XML parsing (server and client side)
- Triple stores (Jena, Sesame, Mulgara, Kowari...)
- Google Earth API (and MS Virtual Earth)
- sealing wax, string, bailing wire, chewing gum...
Common Technologies

- LAMP (Linux + Apache + MySQL + PHP scripting)
- PHP, Perl, Python, shell scripting
- Java (for server side portability) + Tomcat
- XML parsing (server and client side)
- Triple stores (Jena, Sesame, Mulgara, Kowari…)
- Google Earth API (and MS Virtual Earth)
- sealing wax, string, bailing wire, chewing gum…

Open Source, Creative Commons
Common Problems

- Reuse and cross-project leveraging
- Versioning
- Client side scripting and presentation
Common Problems

- Reuse and cross-project leveraging
- Versioning
- Client side scripting and presentation

The *BIG* one!
Applications

● Niche
  – Protein Crystallography
  – Literature Annotation
  – Ontology based services

● General
  – Atlas of Living Australia
  – Water quality monitoring and management
Applications

- **Niche**
  - Protein Crystallography
  - Literature Annotation
  - Ontology based services

- **General**
  - Atlas of Living Australia
  - Water quality monitoring and management

Can agree, or dictate environment
Applications

- **Niche**
  - Protein Crystallography
  - Literature Annotation
  - Ontology based services

- **General**
  - Atlas of Living Australia
  - Water quality monitoring and management

Can agree, or dictate environment

Can’t!
Unfortunate Truth:

Compared to making a light-weight client application operate consistently over all possible host operating systems, browser types, and version combinations, building the Application itself is trivial!
Vannotea

- Browse repositories using inbuilt Internet Explorer
- Simultaneously view and compare multiple media formats
- Annotate files using ontologies; attach local images, pdf's.
- Secure through Shibolleth and XACML
- Search annotations
- Collaborate in real time with Sykpe, Vic, Rat, Jabber
- Record discussions
- Playback with synchronised media reply.
- MS Windows and IE Only
Aus-e-Lit

- Data Integration and search facilities
  - SETIS, NDB, RAAM, PANDORA, ADT, People and Picture Australia, etc
- Empirical Reporting Services
- Compound object authoring, editing and publishing services
- Collaborative Annotation services
Health-e-Waterways

- Assist decision making for scientists, planners, policy makers
- Provide frameworks and services streamlining real-time and near-real-time access to statistical data
- Unite Qld waterway information into a seamless, integrated whole
ARCHE

- Project developed framework services from Monash, James Cook, UQ.
- UQ Components:
  - Metadata Schema Repository (MDSR)
  - Schema driven metadata editor (MDE)
- MDSR stores and versions schemas for use by MDE. Has search and browse interface with ingest facility.
- MDE fully schema driven editor tracks related schema for metadata, including version. Light weight client (EXT based) validated on Linux, MS Windows, and Mac under IE, Firefox, Safari.
**ALA (DIAS-B)**

**Annotation Service**
- Annotation server (“Danno”) implements W3C Annotea protocol, plus OAI-PMH
- Client annotate regions of text or images on *any* web page
- Nothing to install at client
- Javascript client: Firefox, Netscape, Safari, Opera, Chrome, IE
DANNO Stats

- A “cron” job periodically harvests annotations and replies in RDF using OAI-PMH.
- Harvest analysed by a Perl script to generate HTML pages providing activity statistics indicating:
  - Current “hot” topic URLs
  - Most active collaborators
  - Annotation and response counts by page in date order.
Seminar Scorecard

Sharing Data:
Metadata harvesting through OAI-PMH
- ALA (DIAS-B)
- Aus-e-Lit
- Health-e-Waterways

Sharing Ideas:
Collaborative annotations
- ALA (DIAS-B)
- Aus-e-Lit
- Vannotea
Take-away

- The UQ eResearch Lab has more than 10 years temporal experience working with metadata based applications and metadata standards
- We build tools and services that are (mostly) Open Source, configurable, and reusable
- We have a solid history of collaboration with Government and Industry
- We actually enjoy working with metadata!
Key People

- University of Queensland
  - Prof Jane Hunter
    - Principal Investigator
  - Dr Stephen Crawley
    - Senior Research Fellow
  - Mr Ron Chernich
    - Principal Research Fellow
Sponsors

- Australian National Data Service (ANDS)
- Australian Research Collaboration Service (ARCS)
- Australian Collaborative Research Infrastructure Strategy (NCRIS)
- National eResearch Architecture Taskforce (NeAT)
- CSIRO and the ALA
Questions?

- chernich@itee.uq.edu.au
- scrawley@itee.uq.edu.au
- jane@itee.uq.edu.au

More Information and Downloads
http://itee.uq.edu.au/~eresearch