Metadata Stores Program: enabling Discovery and Re-Use

Andrew Treloar
Director of Technology
Overview

- Context
- Why have a metadata store?
- What is a metadata store?
- Why should CAUL care?
- What is the value post-ANDS?
WHY HAVE A METADATA STORE?
Driving and enabling re-use

- Information for discovery
- Information for determination of value
- Information for access
- Information for re-use

- Ideally all of these should be available at discovery time
- And they need to be sourced from somewhere...
From Collections Registry to Discovery
WHAT IS A METADATA STORE?
Metadata Stores Program

- “The Research Metadata Store Infrastructure consists of software to enable the creation, management, and harvesting of a wide range of metadata about data collections and data objects. This includes rich collections description metadata, including links to the data objects (stored elsewhere, such as on the ARCS Data Fabric), data-level descriptions in order to effectively re-use data (scales, field names, variables, calibration settings), and study-level descriptions in order to decide whether to re-use the data or not. This program will develop, configure and make available this metadata infrastructure at research producing institutions.” (ARDC Project Plan, p. 7)
Populating the Collections Registry

**Transport**
- OAI-PMH feed from data sources
- HTTP GET from data sources

**Payload**
- Repository Interchange Format – Collection Schema (RIF-CS)
- Well-formed XML in a known format that ANDS can transform
RIF-CS

- Used to populate Collections Registry
  - name misleading, admittedly
- From which we generate Research Data Australia pages, one for each instance of an ISO 2146 entity
- Defines information about ISO2146 entities
- And relationships between those entities
  - expressed using keys (hence the emphasis on identifiers)
Activity (project): Sound Footings

Activity

View this record in Research Data Australia

Type: project
Key: http://arts.monash.edu.au/music/staff/mkartomi.php+services+200908131020+xs

Originating Source: arrow.monash.edu.au

Group: Monash University

Names:

<table>
<thead>
<tr>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name Parts</td>
</tr>
<tr>
<td>Value</td>
<td>Sound Footings</td>
</tr>
</tbody>
</table>

Related Objects:

Australian Research Council
Key: http://www.arc.gov.au+200908171015+xs

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Description Lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>isFundedBy</td>
<td></td>
<td>en</td>
</tr>
</tbody>
</table>

Kartomi, Margaret J.
Key: http://nla.gov.au/nla.party-493210

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Description Lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasParticipant</td>
<td></td>
<td>en</td>
</tr>
</tbody>
</table>

Monash University, Faculty of Arts, School of Music - Conservatorium
Key: http://arrow.monash.edu.auhdl/1959.1/77627/Party/SchoolOfMusic

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Description Lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasAssociationWith</td>
<td></td>
<td>en</td>
</tr>
</tbody>
</table>

Subjects:

<table>
<thead>
<tr>
<th>Value</th>
<th>Type</th>
<th>Lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>190407</td>
<td>aussrc-for</td>
<td>en</td>
</tr>
</tbody>
</table>
### Kartomi Collection of Traditional Musical Arts in Sumatra

Key:
http://arrow.monash.edu.au/hdl/1959.1/77527

Relations:
- isCollectorOf
- hasAssociationWith

**Description:**
- Description: isCollectorOf
- Description Lang: en

**Additional Details:**
- Monash University, Faculty of Arts, School of Music - Conservatorium
- Australian Archive of Jewish Music Collection
<table>
<thead>
<tr>
<th><strong>Related Objects</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key:</strong> <a href="http://nla.gov.au/nla.party-493210">http://nla.gov.au/nla.party-493210</a></td>
</tr>
<tr>
<td><strong>Description:</strong> hasCollector</td>
</tr>
<tr>
<td><strong>Description Lang:</strong> en</td>
</tr>
<tr>
<td><strong>Monash University, Faculty of Arts, School of Music- Conservatorium</strong></td>
</tr>
<tr>
<td><strong>Key:</strong> <a href="http://arrow.monash.edu.au/hdl/1959.1/77627/Party/SchoolOfMusic">http://arrow.monash.edu.au/hdl/1959.1/77627/Party/SchoolOfMusic</a></td>
</tr>
<tr>
<td><strong>Description:</strong> hasAssociationWith</td>
</tr>
<tr>
<td><strong>Description Lang:</strong> en</td>
</tr>
<tr>
<td><strong>Rechter - Bernard</strong></td>
</tr>
</tbody>
</table>
Metadata store solutions

- Use existing IR
  - Newcastle
- Use existing Research Management solution
  - TBD
- Build something new/integrate
  - ADFI/USQ Fascinator-derivative
  - Griffith/QUT/Unimelb use of VITRO
  - Monash use of MyTARDIS to augment LARDS
  - Monash feed from corporate Oracle applications
- Something else (TBF)
WHY SHOULD CAUL CARE?
Because:

- Libraries care about research outputs
  - and data is now one of these
- Libraries care about metadata
  - and the metadata is critical for discovery and re-use
- Libraries run the IRs
  - which are sometimes the location of data
  - and which are mostly the location of publications
  - and so can contribute significantly to linking these
- Libraries are key stakeholders for institutional information management
WHAT IS THE POINT OF METADATA STORES POST-ANDS?
Post-ANDS...

- Institutions will still need to manage rich metadata about research outputs
- Institutions will still want to track the relationships between Activities, Parties, Services and Collections
- Institutions will still want to publicise their research portfolios
Questions and Links

- ANDS Services site: [http://services.ands.org.au/](http://services.ands.org.au/)

- [andrew.treloar@ands.org.au](mailto:andrew.treloar@ands.org.au)
- [http://andrew.treloar.net/](http://andrew.treloar.net/)
Information for discovery

- Closest thing to catalogue metadata
- Some can be inferred
- Some can be extracted from other systems
- Some has to be entered manually
Information for determination of value

- Helping a researcher decide if they want to re-use
- Based on as much context as possible
  - experimental design
  - researcher
  - research program
  - institution
  - linked publications
  - availability of re-use metadata (see later)
Information for access

- To enable researcher to get to data
- Four possibilities
  - Direct link to open-access data
  - Link to datastore with its own access controls
    - register/login only for open access
    - register/login for restricted access
  - Contact information for how to get data
  - Metadata only (no access)
    - variations in how people feel about this
Information for re-use

- Varies hugely by discipline
- Possible elements
  - variable names and scales (Temp?)
  - instrument settings for experiment
  - calibration values for instrument against reference
  - reagents used
  - rights restrictions