Challenges in building an institutional research data catalogue
DataFinder at the University of Oxford

Sally Rumsey & Neil Jefferies
The Bodleian Libraries

IDCC, Amsterdam, January 2013
Damaro project
Data management rollout for Oxford

DataFinder as the key service for RDM at Oxford
The challenges

Main areas
1. Metadata
2. User interface design
3. Harvesting and interoperability

A solution that serves:
1. A wide variety of data sources
2. A variety of data consumers
Absolute perfection?
## 1. Metadata

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Pragmatic solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandatory fields</td>
<td>Manual deposit only</td>
</tr>
<tr>
<td></td>
<td>+ DataCite core</td>
</tr>
<tr>
<td>Minimum core set</td>
<td></td>
</tr>
<tr>
<td>Context specific</td>
<td></td>
</tr>
<tr>
<td>Tuned to circumstance</td>
<td></td>
</tr>
<tr>
<td>Discipline specificity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Discipline agnostic</td>
</tr>
<tr>
<td>Subject headings</td>
<td>XML files</td>
</tr>
<tr>
<td>Ease of record creation</td>
<td>Common standards</td>
</tr>
<tr>
<td>Additional details</td>
<td>FAST + Keywords</td>
</tr>
<tr>
<td></td>
<td>Auto complete</td>
</tr>
<tr>
<td></td>
<td>Harvest</td>
</tr>
<tr>
<td></td>
<td>Sell the benefits</td>
</tr>
</tbody>
</table>

Aspect Pragmatic solution
### Minimum mandatory metadata (manual)

<table>
<thead>
<tr>
<th>Field to be completed by depositor</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data creator(s) name (given; family)</td>
<td>Other roles available</td>
</tr>
<tr>
<td>Data creator affiliation(s)</td>
<td></td>
</tr>
<tr>
<td>Data owner</td>
<td>For curation and management</td>
</tr>
<tr>
<td>Is the data an output of funded research?</td>
<td>No (no action required)/Yes (requires funder name &amp; grant number)</td>
</tr>
<tr>
<td>Title</td>
<td></td>
</tr>
<tr>
<td>Description/Abstract</td>
<td></td>
</tr>
<tr>
<td>Digital/non-digital</td>
<td>Location required (URL or contact details)</td>
</tr>
<tr>
<td>Publication year</td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>Default University of Oxford</td>
</tr>
<tr>
<td>Earliest date of access</td>
<td>Embargo set here</td>
</tr>
<tr>
<td>Terms and conditions</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Initially FAST headings</td>
</tr>
</tbody>
</table>
2. User interface design

1. Look and feel & branding
2. User journeys
3. Administrative functions
Record status

- **Draft** – user working on record
- **Submitted** – user has submitted record for review
- **Approved** – reviewed submission approved without modification
- **Escalated** – reviewed submission to be checked by other staff due to issues such as commercial or legal agreements, ethics. Note of problem added to admin record.
- **Referred** – more/better information needed before submission can be approved. Submission returned to the submitter with a note of the problem and how to rectify it
- **Rejected** – the admin reviewing the record has decided that there is something fundamentally wrong with it – Reasons for rejection sent to submitter

Diagram:
- User creates new record
  - Draft
    - User edits record
      - User confirms record complete
        - Submitted
          - Reviewer reviews record
            - Senior reviewer
              - Referred
                - Record shunted to catalogue
                  - Explanation sent to submitter; request to re-edit
              - Approved
                - Automated acknowledgement
              - Rejected
                - Reviewer adds comment
                  - Explanation sent to submitter
              - Escalated
                - Record flagged escalated on admin page
Populating DataFinder

Ref: institutional publications repositories
Lack of existing sources
Sell the benefits
  ‘Create once, use many times’
  Citation
  Visibility
  Impact
Mediated service
Questionnaire approach
Aim for ‘good enough’ metadata for citation
Outstanding Challenges

• Obtaining rich, high quality metadata
  – Aim for good enough
  – Encourage rich metadata to generate interest and impact
• Culture change
  – Academics
  – PGRs
• De-duplication
• Sustainability
• Measuring impact
  – Will take time
• Integration with other systems
  • eg Oxford DMPOnline [WIP]
3. Harvesting & interoperability

DataFinder Technical Architecture V1.1

Neil Jefferies, The Bodleian Libraries
http://damaro.oucs.ox.ac.uk/
sally.rumsey@bodleian.ox.ac.uk
neil.jefferies@bodleian.ox.ac.uk