Project StORe
- expectations, a solution
and some predicted impact from
opening up the research data portfolio

Graham Pryor
StORe guide

• Libraries and the research agenda
• The StORe project
  – Objectives
  – Results from a survey
• Support and self-determination
• The challenge for librarians?
Libraries without walls 7
£3,000,000,000 current annual investment – UK Research Councils
Libraries without walls 7

Scholarly publications – the visible research output

Research data – the hidden asset (just melting away?)
The pencil is mightier than the pen
- Robert M Pirsig, *Lila – an inquiry into morals*

• Science as provisional
• Contains a mechanism “whereby new dynamic insight can wipe out old static patterns without wiping out science itself”
• Enables continuous evolutionary growth
The pencil is mightier than the pen

The computer is mightier than the printing press

So, the next stage in the evolutionary growth of science librarians will be…?
Political landscape

• Open access movement – the Berlin, Budapest and other declarations

• Wellcome Trust
  – Funding linked to commitment to deposit

• UK Research Councils – data policies
  – Publicly funded research data a public good
  – Appropriate curation throughout the lifecycle
  – Need for high quality metadata
The StORe mission

- Seamless transport from research data to research publications and vice versa
- Bi-directional links proven in social science e-research but capable of export to other disciplines
Libraries without walls 7

7 domains surveyed for data management and research process

Library origins of project – experience with generic tools and protocols

Survey team: library and domain expertise
New opportunities

• Explore a deeper level of detail
• Supplement published papers
• Validate experiments
• Track the use and improvement of research output
• Identify collaborators
• Confirm completeness of information searches

New risks

• Uncertainty of peer review
• Premature dissemination
• Subversion of scholarly paper
• Scavenging
• Lack of interpretative data
StORe survey

• 85% approved aims as likely to prove advantageous
• Questionable data management practices and a metadata conflict
• Tension between open access and data ownership
Data management

Common practice of storing unique and original research data on the hard drives of PCs and laptops (especially prevalent in chemistry and the biosciences)

“Once removed from the computers I hold the only available copies of most of my data”
Metadata – researchers identified

- The need for improved and universal standards
- A need for different metadata for different phases of the research lifecycle (raw, processed, published data and beyond)
- A clear link between the condition of metadata used and the level of support from information specialists
<table>
<thead>
<tr>
<th>The assignment of metadata</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I decide which terms to use and I assign them:</td>
<td>212</td>
</tr>
<tr>
<td>Research colleague(s) assign metadata on the team's behalf:</td>
<td>55</td>
</tr>
<tr>
<td>Research support staff assign metadata on the team's behalf:</td>
<td>22</td>
</tr>
<tr>
<td>Metadata are assigned by library/information services staff:</td>
<td>4</td>
</tr>
<tr>
<td>Metadata are assigned by the repository administrators:</td>
<td>37</td>
</tr>
<tr>
<td>Metadata are generated automatically:</td>
<td>63</td>
</tr>
<tr>
<td>It is not known who assigns metadata:</td>
<td>68</td>
</tr>
<tr>
<td>Other (please specify):</td>
<td>37</td>
</tr>
</tbody>
</table>
## Libraries without walls 7

<table>
<thead>
<tr>
<th>Types of support used</th>
<th>Academic staff</th>
<th>Research assistant</th>
<th>Post-graduate</th>
<th>Contract researcher</th>
<th>Independent researcher</th>
<th>Other</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documentary support</td>
<td>29</td>
<td>5</td>
<td>16</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>66</td>
</tr>
<tr>
<td>Support by intermediary</td>
<td>16</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>Repository-enabled support</td>
<td>56</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>83</td>
</tr>
<tr>
<td>No support is provided</td>
<td>60</td>
<td>7</td>
<td>23</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>106</td>
</tr>
<tr>
<td>Unknown</td>
<td>31</td>
<td>11</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>78</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>No Answer</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>197</td>
<td>31</td>
<td>83</td>
<td>18</td>
<td>7</td>
<td>41</td>
<td>377</td>
</tr>
<tr>
<td>Assistance by library or IS staff</td>
<td>Academic staff</td>
<td>Research assistant</td>
<td>Post-graduate</td>
<td>Contract Researcher</td>
<td>Independen t researcher</td>
<td>Other</td>
<td>Totals</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>---------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>Documentary (guidance notes etc.)</td>
<td>61</td>
<td>6</td>
<td>24</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>109</td>
</tr>
<tr>
<td>Formal training</td>
<td>38</td>
<td>3</td>
<td>27</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>74</td>
</tr>
<tr>
<td>Online or telephone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>89</td>
</tr>
<tr>
<td>Help structuring searches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>Help in the conduct of searches</td>
<td>20</td>
<td>4</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>42</td>
</tr>
<tr>
<td>Fully managed intermediation (searches and results )</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Unknown</td>
<td>83</td>
<td>19</td>
<td>34</td>
<td>8</td>
<td>3</td>
<td>21</td>
<td>168</td>
</tr>
</tbody>
</table>

“...my university assigned a librarian to our department but I have not used her services...”
The research culture - a broad perspective

• Self-reliance and tradition of individual research endeavour
  – “No assistance required – it’s my responsibility”
  – “I rely on my own skills”

• Protection of research profile or status
  – “Data archives would become the preserve of the scavengers, unable to generate their own research ideas”
  – “The most important factors are those that would benefit my research profile”

• Conflict between a competitive environment and the aspiration to collaborate
  – “The competition for publication is intense and would not be aided by making data freely available”
  – “It should be a requirement for publicly funded research that data be made available”
Factors that *discourage* the sharing of research data:

<table>
<thead>
<tr>
<th>Factor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>The threat of loss of ownership</td>
<td>202</td>
</tr>
<tr>
<td>Risks to an established research niche</td>
<td>104</td>
</tr>
<tr>
<td>Risk of premature broadcast of research findings</td>
<td>235</td>
</tr>
<tr>
<td>Subversion of intellectual property rights, including copyright</td>
<td>163</td>
</tr>
<tr>
<td>Ethical constraints relating to my research</td>
<td>58</td>
</tr>
<tr>
<td>Consideration of data protection and other confidentiality issues</td>
<td>115</td>
</tr>
<tr>
<td>The time/effort required to enable sharing</td>
<td>193</td>
</tr>
<tr>
<td>Risk of diversion from principal objectives through the generation of additional work</td>
<td>144</td>
</tr>
<tr>
<td>Risk to commercialisation opportunities</td>
<td>59</td>
</tr>
<tr>
<td>Increased competition for funding</td>
<td>77</td>
</tr>
</tbody>
</table>
The StORe team concluded that

- Data sharing is a fundamental principle of research, but…
- It is conducted mainly through private networking, with >50% through the personal exchange of e-mails or portable media
- Bureaucratic processes and organisations are anathema to researchers
- Both standards and simplicity are sought for research data management (esp. for metadata)
- Assisted self management is preferred to overt intermediation
Design for the StORe middleware

• Web 2.0 approach, similar to services like Flickr or MySpace, gives control to the researcher
  – Researchers determine which items are public / private
  – Researchers form collaborations with colleagues / ‘friends’
  – Researchers select items for deposit and for publication

• Based on federations of institutional repositories and data archives

• Permanent links created between publications and underlying data

• Simple process for assignment of metadata
  – Searchable metadata assigned at collection level inherited by items within the collection
  – Collection owners add individual items plus minimum additional metadata (e.g. titles)
The purpose of institutional repositories in UK higher education: a repository manager's view


Full text available as:

PDF - Requires Adobe Acrobat Reader or other PDF viewer.

Abstract

Open access publications may be a means of increasing availability of scholarly materials. However, there are other drivers which prompt HE institutions to spend staff time and investment on developing IRs. If the publishing arguments are put to one side, there is a separate and strong case for developing an IR which is justified by a library's functions of collecting, managing and curating materials.

Additional Information: Published 2006 © Elsevier B.V. LSE has developed LSE Research Online so that users may access research output of the School. Copyright and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users
# Source-to-Output Repositories

**Collection:** Madiera

<table>
<thead>
<tr>
<th><strong>Creator</strong></th>
<th>Ken</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
<td>data portal, multilingual thesaurus, online data analysis</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Multilingual Access to Data Infrastructures of the European Research Area</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Contributor</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Identifier</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Relation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Type of Research</strong></td>
<td>Primary from Secondary (Creates new data based on existing data)</td>
</tr>
<tr>
<td><strong>Data Repository</strong></td>
<td>UKDA</td>
</tr>
<tr>
<td><strong>Study</strong></td>
<td>4995</td>
</tr>
<tr>
<td><strong>Visibility</strong></td>
<td>Private</td>
</tr>
<tr>
<td><strong>Owner</strong></td>
<td>Ken Miller (forum)</td>
</tr>
<tr>
<td><strong>Contributors</strong></td>
<td></td>
</tr>
</tbody>
</table>
Item: Business Analysis
Collection: Edinburgh

Dublin Core

Edit

* Denotes required fields:

Title: Business Analysis
Creator: Ken Hills
*Subject: ad
*Subject: presentation
Description: A collaborative collection that I might use in the pilot demonstration in Edinburgh
Publisher: Ken Hills
Contributor: 
Date: 07/11/2006
Type: 
Format: 
Identifier: 
Source: 
Language: English
Relation: 
Coverage: 
Rights: 
Type of Research: Secondary Analysis Based on existing data
Study: 

Add Subject Tags:
- demo
- demonstration
- email
- health
- subject
- pilot
- preservation
- data publications
- image
- subject

Add Selected
Roles, Rights, Responsibilities, Relationships - 1

- Institutional Repositories tend to be managed by ‘generalists’ located in libraries and information services, and usually lacking data skills

- Subject repositories are staffed by domain experts with well-developed data handling skills

- Data sets tend to be unique
  - are deposited without multiple copies
  - are not self-describing and may
  - require interpretation, particular competencies and domain knowledge

- Representation information, which may include software programs, is required to assist with interpretation

- Closer partnerships between institutional and subject repository staff will be essential to building engagement with the designated community

Dealing With Data, Dr Liz Lyon, UKOLN, June 2007
Institutional repositories have a fundamental argument for sustainability because of their position within an institution.

They are frequently located in departments with a persistent service orientation - i.e. libraries or information services - whose main role is the management of information.

Many scholars care primarily about their discipline or domain, before their institution.

Scholars who have hybrid skills in domain science and data/information handling will be a valuable resource in the future.

Dealing With Data, Dr Liz Lyon, UKOLN, June 2007
Libraries without walls 7

The next stage in the evolutionary growth of librarians...

Re-emergence of the scholar librarian?
Conclusion

Thank you