Partnering for research data

Dr Liz Lyon, Associate Director, UK Digital Curation Centre
Director, UKOLN, University of Bath, UK

2nd LIBER International Workshop on Digital Preservation,
Florence, May 2012

This work is licensed under a Creative Commons Licence
Attribution-ShareAlike 2.0
Data...
Research data stakeholders

- Who are they?
- Roles and responsibilities?
- Infrastructure requirements, tools, and skills?
- Gaps and opportunities?
Partners: European....
Partners: global....
Stakeholders within the University

- Roles (7 listed)
- Responsibilities
- Requirements
- Relationships

Institutional partners

Liz Lyon, Informatics Transform, IJDC Current Issue, 2012
1. Director IS/CIO/University Librarian
2. Data librarians /data scientist /liaison/subject/faculty librarians
3. Repository managers
4. IT/Computing Services
5. Research Support/Innovation Office
6. Doctoral Training Centres
7. PVC Research
8. + Public Engagement Office

Liz Lyon, Informatics Transform, IJDC Current Issue, 2012
<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibilities</th>
<th>Requirements</th>
<th>Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Director Information Services / CIO</td>
<td>To lead and co-ordinate data informatics support</td>
<td>Appropriate LIS structure in place</td>
<td>PVC Research, Deans, Associate Deans, Faculty/School Directors of Research, IT Director, Director Research Support</td>
</tr>
<tr>
<td>University Librarian</td>
<td></td>
<td>Library staff with data informatics &amp; research data management skills</td>
<td>Other key institutional stakeholders</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Institutional repository with content links to underlying research data</td>
<td>Open Access Publishers</td>
</tr>
<tr>
<td>Data librarian / Data scientist / Liaison</td>
<td>To deliver expert data informatics advice and guidance to research staff</td>
<td>Knowledge of data management planning and data audit and assessment tools</td>
<td>DTCs, post-grads, PIs</td>
</tr>
<tr>
<td>/Subject / Faculty Librarian</td>
<td>To facilitate access to datasets for PIs, research staff, postgraduate and</td>
<td>Knowledge of selection and appraisal, metadata standards and schema, data</td>
<td>DCC</td>
</tr>
<tr>
<td></td>
<td>undergraduate students</td>
<td>formats, domain ontologies, identifiers, data citation, data licensing</td>
<td>DataCite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knowledge of appropriate disciplinary data centres,</td>
<td>Data centre staff</td>
</tr>
</tbody>
</table>

Full mapping: Informatics Transform, IJDC Current issue, 2012
<table>
<thead>
<tr>
<th>Role</th>
<th>Responsibilities</th>
<th>Requirements</th>
<th>Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repository managers</td>
<td>To ensure research papers have persistent links to underlying research data</td>
<td>Knowledge of persistent identification mechanisms and publisher requirements</td>
<td>Data librarians / Data scientists / Liaison / Subject / Faculty Librarians</td>
</tr>
<tr>
<td>IT / Computing Services</td>
<td>To provide data storage infrastructure and guidance</td>
<td>Knowledge of data storage options including cloud-based services</td>
<td>EduServ data centre. Cloud service providers</td>
</tr>
<tr>
<td>Research &amp; Development Support</td>
<td>To provide RIM/CRIS capability for research outputs</td>
<td>Provision for non-textual outputs such as datasets, software and program code, gene sequences, models</td>
<td>Research funding bodies</td>
</tr>
<tr>
<td>Office / Research &amp; Innovation</td>
<td></td>
<td></td>
<td>Data scientists / Liaison / Subject / Faculty Librarians</td>
</tr>
<tr>
<td>Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role</td>
<td>Responsibilities</td>
<td>Requirements</td>
<td>Relationships</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Faculty Doctoral Training Centres</td>
<td>To supply training to new-entrant researchers and PIs</td>
<td>Knowledge of data management planning and data audit and assessment tools</td>
<td>Deans &amp; Associate Deans, PIs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Training programmes and modules</td>
<td>Data librarian / Data scientist / Liaison / Subject / Faculty Librarians</td>
</tr>
<tr>
<td>PVC Research</td>
<td>To develop institutional research policy and code of practice</td>
<td>Understanding of data management compliance implications, risks including legal and ethical issues, and sustainability challenges</td>
<td>Deans &amp; Associate Deans</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Key service directors</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Research &amp; Development Support Office / Research &amp; Innovation Services</td>
</tr>
<tr>
<td>Public Engagement Unit</td>
<td>To facilitate citizen participation in the research process</td>
<td>Understanding of open science methodologies and infrastructure</td>
<td>PVC Research Director, Communications Deans &amp; Associate Deans, PIs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The Media</td>
</tr>
</tbody>
</table>
Research360@Bath

• Partnership approach
  • UKOLN-DCC
  • Library
  • IT Services
  • Research Support Office
  • Doctoral Training Centres
  • Data Scientist

http://blogs.bath.ac.uk/research360/
Infrastructure requirements:
Delivering RDM Services

(Providing tools & support)
Understanding Data Requirements

If research data lies at the heart of your organisation, you need to know that you have adequate infrastructure, staff skills and resources, and senior management support in place to ensure that your data is effectively managed for validation, reuse and evidential purposes.

**CARDIO enables you to:**

- Collaboratively assess data management requirements, activity, and capacity at your institution
- Build consensus between data creators, information managers and service providers
- Identify practical goals for improvement in data management provision and support;
- Identify operational inefficiencies and opportunities for cost saving;
- Make a compelling case to senior managers for investment in data management support
Data management plans
• Advocacy & Training
  • Informatics: disciplinary metadata schema, standards, formats, identifiers, ontologies
  • Storage: file-store, cloud, data centres, funder policy
  • Access: embargoes, FOI
What data to keep

How to Cite Datasets and Link to Publications
Alex Ball (DCC) and Monica Duke (DCC)

How to Appraise & Select Research Data for Curation
Angus Whyte (DCC) and Andrew Wilson (ANDS)
Data Licensing

Bespoke licences
Standard licences
Multiple licensing
Licence mechanisms
Tools to track impact

Uncover the invisible impact of research.

Create a collection of research objects you want to track. We'll provide you a report of the total impact of this collection. You can peruse a sample report or check out the most recently shared reports.

Collect research objects

Paste object IDs,
Add one DOI, PubMed ID, URL, or other supported identifier per line:
- 10.1234/journal.pabi.123456
- 20345678
- ZABK
- G12345
- 12.3456/dryad.1234
- https://www.snowbeetles.info/research/lab/notebook
- https://www.slideshare.net/phylogenomics/eisenall-hands

...or pull object IDs from existing collections.
- Mendeley profiles
- Mendeley groups
- Slideshare accounts
- Dryad dataset authors
- PubMed grants
- GitHub users
- GitHub organizations

Create report

Name your collection:
my collection

... or fetch a quick collection based on your Mendeley contacts and public groups »

http://total-impact.org/
Data informatics capacity & capability

(Acquiring the skills....)
Sheila Corrall: Libraries, Librarians and Data
Many action exemplars

RLUK/Mary Auckland: Reskilling for Research
9 areas are skill gaps for subject librarians

2012: Libraries in review
<table>
<thead>
<tr>
<th>Skill gap</th>
<th>2-5 years</th>
<th>Now</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preserving research outputs</td>
<td>49%</td>
<td>10%</td>
</tr>
<tr>
<td>Data management &amp; curation</td>
<td>48%</td>
<td>16%</td>
</tr>
<tr>
<td>Comply with funder mandates</td>
<td>40%</td>
<td>16%</td>
</tr>
<tr>
<td>Data manipulation tools</td>
<td>34%</td>
<td>7%</td>
</tr>
<tr>
<td>Data mining</td>
<td>33%</td>
<td>3%</td>
</tr>
<tr>
<td>Metadata</td>
<td>29%</td>
<td>10%</td>
</tr>
<tr>
<td>Preservation of project records</td>
<td>24%</td>
<td>3%</td>
</tr>
<tr>
<td>Sources of research funding</td>
<td>21%</td>
<td>8%</td>
</tr>
<tr>
<td>Metadata schema, discipline standards, practices</td>
<td>16%</td>
<td>2%</td>
</tr>
</tbody>
</table>

*Data from RLUK/Mary Auckland: Reskilling for Research 2012*
“Very few librarians are likely to have specialist scientific or medical knowledge - if you train as a research scientist or a medic, you probably won’t become a librarian.”
<table>
<thead>
<tr>
<th>Position</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Data Librarian</td>
<td>Stanford</td>
</tr>
<tr>
<td>Data Management Librarian</td>
<td>Oregon State</td>
</tr>
<tr>
<td>Social Sciences Data Librarian</td>
<td>Brown</td>
</tr>
<tr>
<td>Data Curation Librarian</td>
<td>Northeastern</td>
</tr>
<tr>
<td>Data Librarian</td>
<td>New South Wales</td>
</tr>
<tr>
<td>Research Data Management Co-ordinator</td>
<td>Sydney</td>
</tr>
<tr>
<td>Research Data &amp; Digital Curation Officer</td>
<td>Cambridge</td>
</tr>
<tr>
<td>Data Services Librarian</td>
<td>Iowa</td>
</tr>
<tr>
<td>Data Analyst</td>
<td>ANDS</td>
</tr>
<tr>
<td>Institutional Data Scientist</td>
<td>Bath</td>
</tr>
</tbody>
</table>
Implications of “Big Data” and data science for organisations in all sectors

Predicts a shortage of 190,000 data scientists by 2019

http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation
Gaps? Opportunities??

1. Define core components of data informatics

- Metadata (discovery, preservation)
- Domain ontologies
- Visualisation e.g. VisTrails
- Workflow e.g. Taverna
- Analysis e.g. R

*Lyon, Informatics Transform, IJDC 2012*
Gaps? Opportunities??

2. Analyse LIS entry qualifications & increase STEM entrants

Target
- Biologists
- Chemists
- Mathematicians

Lyon, Informatics Transform, IJDC 2012
Gaps? Opportunities??

3. International Data Informatics Working Group to explore promotion, recognition & reward

- Global awareness campaign
- Career incentives
- Benchmark good practice

*Lyon, Informatics Transform, IJDC 2012*
“The ability to take data - to be able to understand it, to process it, to extract value from it, to visualise it, to communicate it - that’s going to be a hugely important skill in the next decades.”

Hal Varian, Chief Economist, Google
8th International Digital Curation Conference, Amsterdam, 14-16 January 2013
Thank you!

Informatics Transform article
http://www.ijdc.net/index.php/ijdc/article/view/210

Slides
http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/presentations.html

DCC http://www.dcc.ac.uk