Continuity and change
Opportunities and challenges for the future of
research libraries in a data-intensive age

Michael Day
Digital Curation Centre
UKOLN, University of Bath, UK
m.day@ukoln.ac.uk

5th Joint Conference of Slovene Special and Academic Libraries,
Ljubljana, Slovenia, 17-18 October 2012
Pulteney Bridge, Bath

http://www.flickr.com/photos/13706945@N00/7756014816/
Presentation outline

- Changing research library roles
- **Research Data Management**
  - Changes in policies
  - Changes in research practice
  - Open Access and Open Science
- The skills gap
- Potential roles for the library in supporting RDM
- Some concluding thoughts

  - Presentation slides can be found at: http://www.slideshare.net/michaelday/continuity-change-14758361
## Research library roles

<table>
<thead>
<tr>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection development and acquisitions</td>
</tr>
<tr>
<td>Cataloguing and knowledge organisation</td>
</tr>
<tr>
<td>Circulation</td>
</tr>
<tr>
<td>Reference work; current awareness</td>
</tr>
<tr>
<td>Preservation and conservation</td>
</tr>
</tbody>
</table>
# Research library roles

<table>
<thead>
<tr>
<th>Traditional</th>
<th>Emerging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection development and acquisitions</td>
<td>Content creation (e.g. digitisation); license negotiation; dealing with new forms of content</td>
</tr>
<tr>
<td>Cataloguing and knowledge organisation</td>
<td>Resource description and metadata; linked data</td>
</tr>
<tr>
<td>Circulation</td>
<td>Empowering users, e.g. training; awareness of institutional contexts</td>
</tr>
<tr>
<td>Reference work; current awareness</td>
<td></td>
</tr>
<tr>
<td>Preservation and conservation</td>
<td>Digital preservation and sustainability, repositories</td>
</tr>
</tbody>
</table>
Research data management (1)

- RDM is rapidly emerging as an area of great interest to research libraries in the UK
  - Policy drivers are encouraging institutions to deal with this issue now
  - Institutions look to their libraries for advice on this
  - Libraries keen to clarify their role in supporting RDM, e.g. building on their experiences with implementing Institutional Repository infrastructures
Research data management (2)

- **Why has RDM become important now?**
  - Changes in research practice (e.g., data-intensive research)
  - Changes in the expectations and policies of funding bodies
  - Changes in the culture of research (e.g. Open Science)
Data-intensive research

- Jim Gray’s “Fourth Paradigm”
- **Difficult to define, but (broadly speaking) involves:**
  - Research involving large amounts of data
  - Data is combined from multiple sources, across multiple disciplines
  - Data requiring significant processing (computational analysis)
- **Becoming increasingly embedded in research practice**
  - Integral for many ‘big science’ disciplines
  - Now influencing “long-tail sciences,” the humanities and social sciences

www.ukoln.ac.uk

A centre of expertise in digital information management
Research funding bodies

- **UK Research Councils**
  - Already help fund some data archives, e.g.:
    - Archaeology Data Service, European Bioinformatics Institute, the NERC data centres, UK Data Archive
  - Research Councils UK (RCUK) Common Principles on Data Policy
    - Recognises that data are a critical output of the research process
    - [http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx](http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx)
RCUK Principles (in a nutshell)

- Publicly funded research data should be made openly available
- Data with acknowledged long-term value should be preserved and remain accessible and usable for future research
- Sufficient metadata should be recorded to enable other researchers to find and understand the research to enable re-use; published results should always include information on how to access the supporting data
- Recognition that there may be legal, ethical and commercial constraints
- Recognition that researchers may need privileged use of data for a limited period
- All users of research data should acknowledge their sources
- Appropriate to use public funds to support MRD
EPSRC Policy Framework (1)

- Engineering and Physical Sciences Research Council (EPSRC) Policy Framework on Research Data (2011)
  - Required a “Roadmap” by May 2012; compliance by May 2015
  - EPSRC Expectations have been a major influence in getting RDM onto the agenda of senior management in UK universities
  - Institutions with significant EPSRC funding could not afford to ignore this

- http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx
EPSRC Policy Framework (2)

- **Selected points:**
  - Appropriate metadata (including unique IDs) to be made freely available on the Internet within 12 months of data generation.
  - Data not generated in digital format should be stored in a manner that facilitates it being shared.
  - Data should be securely preserved for a minimum of 10 years after privileged access expires or the last date access was requested by a third party.
  - There should be adequate resources from existing funding streams.
  - EPSRC will monitor progress and compliance, reserving the right to impose “appropriate sanctions”.

www.ukoln.ac.uk

A centre of expertise in digital information management
Research360@Bath

- New institutional data scientist role
- Addresses EPSRC expectations (published)
- Doctoral Training Centre hubs
- Faculty-Industry focus
- Faculty cascade model
- Multi-team approach

http://blogs.bath.ac.uk/research360/

www.ukoln.ac.uk

A centre of expertise in digital information management
Open Access

- Open Access
  - Growing expectation from governments and funding bodies that outputs from publicly-funded research should be freely available
  - Supported by funding body and institutional mandates
  - Two main approaches to date:
    - Institutional Repositories, typically (86%) hosted and supported by libraries; some subject-based repositories
    - Steady growth in the number of Open Access journals, often funded by article processing charges
The Finch Report

- Working Group on Expanding Access to Published Research Findings
- Report: “Accessibility, sustainability, excellence: how to expand access to research publications” (June 2012)
  - Noted that the rate of deposit of papers in repositories so far “has been disappointing” (p. 6)
  - Recommendations included:
    - Proposed accelerating the transition to Open Access through publication in OA or hybrid journals
    - Recommended developing mechanisms to enable universities to pay article processing charges
    - Repositories to focus on reports, working papers, theses and dissertations
RCUK Open Access policy

- Research Councils UK Policy on Access to Research Outputs (July 2012)
  - Based on the recommendations of the Finch Report
  - Peer reviewed research papers which result from research that is wholly or partially funded by the Research Councils:
    1. must be published in journals which are compliant with Research Council policy on Open Access
    2. must include details of the funding that supported the research, and a statement on how the underlying research materials – such as data, samples or models – can be accessed
Science as an open enterprise (1)

- OA to research outputs part of a wider trend of opening up access to research practice more generally
- Royal Society report “Science as an open enterprise” (June 2012)
  - Focused on the potential of the “data deluge”
  - Produced a very thorough analysis of the current state-of-play on research data management in the UK
Science as an open enterprise (2)

- Recommendation 1: “Scientists should communicate the data they collect and the models they create, to allow free and open access, and in ways that are intelligible, assessable and useable for other specialists in the same or linked fields wherever they are in the world. Where data justify it, scientists should make them available in an appropriate data repository.”
Science as an open enterprise (3)

- Report identified four tiers of data management:
  - Tier 1 – major international initiatives
  - Tier 2 – national data centres
  - Tier 3 – institutions (universities and research centres)
  - Tier 4 – individual researchers or research groups

- Concern about Tier 3 institutional strategies
  - Questions what responsibility institutions should have in supporting the data curation needs of their researchers and in curating the data produced by researchers
  - Implications for libraries … “
Science as an open enterprise (4)

- From Section 4.2.1
- “A particular dilemma for universities is to determine the role of their science libraries in a digital age. In the majority of cases (86%), libraries have led responsibility for the university repository. The traditional role of the library has been as a repository of data, information and knowledge and a source of expertise in helping scholars access them. That role remains, but in a digital age, the processes and the skills that are required to fulfil the same function are fundamentally different.”
Reskilling for research (1)

- Research Libraries UK report
- Written by Mary Auckland, former Director of Library and Learning Resources at the University of the Arts London;

- Available:
  http://www.rluk.ac.uk/content/re-skilling-research
Reskilling for research (2)

- Auckland identified 9 key areas with skill gaps for subject librarians:
  - Ability to advise on **preserving research outputs**
  - Knowledge to advise on **data management and curation**, including ingest, discovery, access, dissemination, preservation, and portability
  - Knowledge to support researchers in **complying with the various mandates of funders**, including open access requirements
  - Knowledge to advise on potential **data manipulation tools** used in the discipline/subject
  - Knowledge to advise on **data mining**
Reskilling for research (3)

- **Key areas (continued)**
  - Knowledge to advocate, and advise on, the use of **metadata**
  - Ability to advise on the **preservation of project records** e.g. correspondence
  - Knowledge of **sources of research funding** to assist researchers to identify potential funders
  - Skills to develop **metadata schema, and advise on discipline/subject standards and practices**, for individual research projects
<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>

Source: Auckland, *Reskilling for research* (2012)
Potential library roles (1)

- Leadership, co-ordinating action across the institution (and beyond)
  - Researchers – both as creators and users of data, PIs (have specific roles on grants), data scientists
  - Administration – research support offices, records managers, FOI officials
  - Central services – e.g., computing services

- Training
  - Responding to researchers’ training needs
  - Continuing professional development for information professionals

www.ukoln.ac.uk

A centre of expertise in digital information management
Potential library roles (2)

- **Policy development**
  - Approval of policies by senior management is just the start; policies need to be embedded in research practice and responsive to changing requirements

- **Data management planning**
  - DCC How-to Develop a Data Management Plan guide: http://www.dcc.ac.uk/resources/how-guides/develop-data-plan
  - DMP online: http://www.dcc.ac.uk/dmponline
Potential library roles (3)

• Requirements gathering
  – Identifying researchers’ data requirements
  – Developing a shared understanding of what needs to be done (e.g., identifying where data exist, its form and scale, any existing retention requirements)
  – Identifying good practice within the institution (and the opposite)
  – Methods: surveys, focus groups, case studies, joint R&D projects, assessment tools, e.g. DCC’s Digital Asset Framework (DAF)
Potential library roles (4)

- **Assessing institutional preparedness**
  - Identifying institutional stakeholders, existing data support services, gaps
  - Benchmarking and planning for the future
  - Skills audits
  - DCC CARDIO (Collaborative Assessment of Research Data Infrastructure and Objectives) benchmarking tool
  - [http://cardio.dcc.ac.uk/](http://cardio.dcc.ac.uk/)
Potential library roles (5)

- **Identifying motivations and benefits**
  - For researchers, support services, the institution

- **Identifying risks**
  - Data loss (institution, research group, individual)
  - Increased costs (lack of planning, service inefficiency, data loss)
  - Legal compliance (research funder, H&S, ethics, FoI)
  - Reputation (institution, unit, individual)

- **Identifying costs**
  - Keeping Research Data Safe (KRDS) toolkit

www.ukoln.ac.uk

A centre of expertise in digital information management
Potential library roles (6)

- **RDM infrastructures**
  - Integrating where possible with existing services, e.g. IR, CRIS, VRE, HPC, cloud services, social media, etc.
  - Storage choices – no one-size-fits-all solution, e.g. Bristol’s BluePeta petascale storage facility, Bath’s X-Drive approach, cloud approaches

- **Metadata**
  - Data documentation and metadata – layered approaches: top-level discovery (core metadata, collection/experiment-level?), role of standards like DCMI, CERIF, DDI, DOI, ORCID, etc.
Potential library roles (7)

- **Selection and appraisal**
  - Deciding what needs to be kept and for how long
  - DCC guide: http://www.dcc.ac.uk/resources/how-guides/appraise-select-data
Potential library roles (8)

• Data citation
  – Important to link publications to data (and vice versa); a requirement of some journals
  – Increases citations – of both data & publication
  – Data citation potentially increases reuse (hence value)
  – DCC guide: http://www.dcc.ac.uk/resources/how-guides/cite-datasets
Potential library roles (9)

• Analysing tools that 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:
10.1071/journal.publ.1000061
2002622
30050
1.0061/droyed.129
http://www.ceridostiger.info/research/
lab-notebook
http://www.slideshare.net/phylogemomics
/kewa23-hands

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

Name your collection:

my collection

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

Something missing on import? See a list of current limitations.
DCC Institutional engagements

Assess needs

- Workflow assessment
- DAF & CARDIO assessments
- Advocacy with senior management

DCC support team

- Institutional data catalogues
- Pilot RDM tools
- Guidance and training
- RDM policy development
- Customised Data Management Plans

Develop support and services

Make the case

- ...and support policy implementation

www.ukoln.ac.uk

A centre of expertise in digital information management
Continuity and change …

<table>
<thead>
<tr>
<th>Traditional</th>
<th>Emerging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection development and acquisitions</td>
<td>Content creation (e.g. digitisation); license negotiation; dealing with new forms of content</td>
</tr>
<tr>
<td>Cataloguing and knowledge organisation</td>
<td>Resource description and metadata; linked data</td>
</tr>
<tr>
<td>Circulation</td>
<td>Empowering users, e.g. training; awareness of institutional contexts</td>
</tr>
<tr>
<td>Reference work; current awareness</td>
<td></td>
</tr>
<tr>
<td>Preservation and conservation</td>
<td>Digital preservation and sustainability, repositories</td>
</tr>
</tbody>
</table>
Hvala lepa | Thank you!
Acknowledgments

• The Digital Curation Centre (DCC) is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK's higher education research community. The DCC is funded by JISC.

• More information is available from: http://www.dcc.ac.uk/

• UKOLN receives support from JISC and the University of Bath, where it is based.

• More information is available from: http://www.ukoln.ac.uk/