Roadmaps, Roles & Re-engineering: Developing Data Informatics Capability in Libraries

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

LIBER Conference Munich, June 2013

This work is licensed under a Creative Commons Licence Attribution-ShareAlike 2.0
Direction of travel

• Data headlines 2013
• University as a data publisher
• What is the Library data offer?
• Future look towards re-invention
Community Capability Model WG

The RDA Community Capability Model (CCM) Working Group (WG) collects, validates and publishes a range of data-centric “capability profiles” to enhance inter- and intra-domain interoperability and catalyse RDA data-sharing goals.

Open Research Data ... breaking down barriers - RDA Second Plenary Meeting

16-18 September 2013 - National Academy of Sciences, Washington DC, US. Data sharing offers important benefits for scientific progress and advancement of knowledge. However, several limitations and barriers in the general adoption of data sharing are still in place.
UK government supports open data

Independent review of public sector information May 2013
Endorses OA

Open Data Charter

Policy Paper

18 June 2013

Policy paper

G8 Open Data Charter and Technical Annex

Published 18 June 2013

Contents
1. Principle 1: Open Data by Default
2. Principle 2: Quality and Quantity
3. Principle 3: Usable by All
4. Principle 4: Releasing Data for Improved Governance
5. Principle 5: Releasing Data for Innovation
6. Technical annex
Open access to scientific information: a way to get there

Open access (OA) = online access at no charge to the user
- to peer-reviewed scientific publications
- to research data

Policy implications for institutions

Roadmaps

Planning frameworks
Policy Framework on Research Data 2011
- Principles and Expectations
- http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx
### Alignment with EPSRC Expectations

#### Roles and Responsibilities:
Who’s responsible

#### Objectives:
Where we need to be

#### Where we are now

#### Actions:
How we’re going to get there

#### Milestones:
When it will be done by

---

**EPSRC Expectations**

1. Research organisations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data.

Research undertaken at the University already complies with all relevant legislation, including data protection, Freedom of Information, copyright, ethical and environmental information, regulations. The Research Development Support Office and the Freedom of Information Coordinator provide bespoke guidance to researchers who have reason to withhold their research data.

---

**Proposed Activities**

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Actions</th>
<th>Milestones</th>
<th>Roles &amp; Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Develop the data management skills and knowledge of Bath researchers</td>
<td>• Initiate and coordinate a program of research data management training</td>
<td>• Produce skills development in research data management</td>
<td>• Library/IRIS</td>
</tr>
<tr>
<td>2. Enhance the skills and knowledge of key professional services staff</td>
<td>• Work with Faculty, RDG and Bath Graduate Schools to develop and deliver training in research data management skills</td>
<td>• Develop skills development in research data management</td>
<td>• RDSG</td>
</tr>
<tr>
<td>3. Enhance the skills and knowledge of key professional services staff</td>
<td>• Seek opportunities for professional services staff to network and participate in external networks</td>
<td>• Enhance research data management policy and professional support services</td>
<td>• Staff Development Unit</td>
</tr>
</tbody>
</table>

---

**Roles and Responsibilities**

- **Library/IRIS**
- **RDSG**
- **Staff Development Unit**
- **Data Scientist**

---

- Bath Research Data Sustainability Group (RDSG)
- University stakeholders represented by Heads of Service/Assoc Deans

---

[http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf](http://www.bath.ac.uk/rdso/University-of-Bath-Roadmap-for-EPSRC.pdf)
Roles

New Library data roles too 😊

Transition from pilot RDM projects to sustainable and integrated data services

Articulate business case for RDM investment….. but first…..
210 respondents (3.5% response rate): PIs, ROs, postgrads

Some preliminary findings:
- Most have not had to produce a data management plan (81%)
- Much data is confidential, anonymised, under non-disclosure agreements, commercially sensitive, DPA, encrypted
- Data is in non-digital form: lab notebooks, interviews
- Researchers store data on Univ Bath shared filestore 😊
- They also use Dropbox, USB sticks, home computers 😞
- Data loss: accidental deletion, hardware failure, obsolescence
- Open data is not the norm – often shared informally
- Lack of recognition for data sharing and reuse is an issue

Know the state-of-the-nation

These findings mirror most/all institutional data surveys...
Understand the costs

- A rather grey area…..
- UK Research funder grants
- Direct costs (during lifetime of grant)
- Funder expectations
- Long-term RDM investment an institutional responsibility
Bath RDM Business Case

• Worked up with Neil Beagrie
• Presented to V-C Group in March 2013
• Positive outcome: 2 f/t permanent posts in Library
• Data Scientist & Technical Data Officer

Benefits of investment
Risks if none
Options & levels
Recommendations
Re-engineering...

From 300 year old print-based traditions....

....to data informatics services &
....data-savvy staff
Positioning the university as a data publisher

- Takes responsibility for its data products
- HEI stakeholders described in 2012 (Informatics Transform)
- Advocacy gap
- Infrastructure gap
- Reputation?
- Risk?
Data Integrity

I write to alert the organic chemistry community to a serious problem related to the integrity of data being submitted for review and publication by *Organic Letters* and to outline steps that the Journal is taking to address this concern. Recently, with the addition of a Data Analyst to our staff, *Organic Letters* has begun checking the submitted Supporting Information more closely. As a result of this increased scrutiny, we have discovered several instances where reported spectra had been edited to remove evidence of impurities.

Such acts of data manipulation are unacceptable. Even if the experimental yields and
Advocacy, Awareness, Training

- Researcher awareness is low
- Advocacy needed (like OA)
- Change behaviour to “open by default” (G8)
- Open data literacy (G8)
- RDM training: Research360 & Doctoral Training Centres
- R360 RDM online learning module
Data Management Plans

Analysed requirements

Developed a Checklist

Provided tools & guidance

DMPonline
The DCC Data Management Planning Tool

Links to all DMP resources via http://www.dcc.ac.uk/resources/data-management-plans

2007

2011
2013 DMP developments

• New radically simplified DCC DMP Checklist with funder questions – consultation and revision
• New DMP taxonomy

• DMPonline tool Vs3 road-tested with researchers
• Four User Roles with Use Cases
• Vs 4 beta version in September

• Taxonomy as basis for international DMP Profile (CERIF-compliant) development by CASRAI WG
Bath institutional DMP guidance

Writing your data management plan

University of Bath: DMP template guidance

Introduction
- What is data?
- Why plan for data management
- How to approach data management planning

Completing the plan
- Administrative details
- Data creation/collection
- Data management, documentation and preservation
- Data security
- Data archival and preservation
- Data publication and access
- Roles, responsibilities and resourcing

Further resources

DATA MANAGEMENT PLAN

<table>
<thead>
<tr>
<th>Principal investigator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-investigator(s)</td>
</tr>
<tr>
<td>Project title</td>
</tr>
<tr>
<td>Funder(s)</td>
</tr>
<tr>
<td>Start date</td>
</tr>
<tr>
<td>End date</td>
</tr>
</tbody>
</table>

Many questions include example options for answers - use or delete as appropriate, or replace with your own response.

Data creation/collection

What existing sources of data will be used?
- Own/group previous research
- Academic collaborators
- Commercial collaborators

What are the characteristics of the data?
- Quantitative
- Images
Institutional data publication services

- Data repository: curation
- Data catalogue: discovery
- Data citation: Amsterdam Manifesto, identifiers, DOIs, ORCID, data licences
- New actionable formats: data papers, data journals, + data peer review....
- Data metrics: ImpactStory (UK REF2020?)
Choice of RDM training materials for librarians

Do-It-Yourself Research Data Management Training Kit for Librarians

Up-skilling for data

http://dataintelligence.3tu.nl/en/home/
Developing data capability in professional services

http://immersiveinformatics.org

**Immersive Informatics** pilot at University of Bath
Co-developed with University of Melbourne
July 2013

10 modules (OER)

Day release

**Immersive** data sessions in labs

Co-curate dataset

Keep “data diary”
Futures...

Some closing provocations...
Every research project has a DMP
Library routinely provides DMP support
DMPs are harmonised with common core elements
DMPs are peer reviewed, stored, analysed and mined for data intelligence by the Library
DMP analytics inform institutional planning & validate research infrastructure investments
Data is mission-critical for Libraries

1. Library delivers core RDM services
2. Regional partnerships with collaborative and shared data infrastructure services
3. Demonstrate cost-benefits and efficiency savings
4. Data scientist, data librarian, data analyst teams
5. Transformational data service delivery modes
Librarians in the Lab

1. Co-locate data informatics support at the research coalface fully-integrated in research workflow
2. Data librarians listed as “co-authors” in citations with due attribution and credit
3. Recruit more LIS staff from STEM disciplines for data analysis, data visualisation, data stories
4. Data literacy in core LIS education curriculum
5. Re-invent the Library & re-invent LIS careers…
Thank you!

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

ImmersiveInformatics
http://immersiveinformatics.org

Research360 Project at University of Bath
http://blogs.bath.ac.uk/research360/

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