Data: opportunities for libraries

Kevin Ashley
Digital Curation Centre
www.dcc.ac.uk
@kevingashley
Kevin.ashley@ed.ac.uk

Reusable with attribution: CC-BY

The DCC is supported by Jisc
A summary

• A recap of why anyone cares about research data
• A focus on research data: library roles, institutional roles
• A nod to other data-related roles for libraries
• Note absence of BIG
  – A large collection of small data is big data
• But first, an apology
I am an intruder

That's not why I'm here
My home – the DCC

• Mission – to increase capability and capacity for research data services in UK institutions
• Not just a UK problem – an international one
• Training, shared services, guidance, policy, standards, futures
What is data curation?

- “Maintaining, preserving and adding value to research data throughout its lifecycle”
- More than preservation:
  - Active management – dealing with change
- Less than preservation:
  - Lifecycle sometimes involves destruction
- Sometimes, not always, about sharing, publication or citation
Why care?

• Data is expensive – an investment
• Reuse:
  – More research
  – Teaching & Learning
  – Planning
• Impact – with or without publication
• Accountability
• Legal & regulatory requirements
Why does this matter?

• Research quality
  – How close can we get to the truth?

• Research speed
  – How quickly can we get to the truth?

• Research finance
  – How much does the truth cost?

• Improving one or more of these is of interest to all actors:
  • Researchers as data creators
  • Researchers as data reusers
  • Research institutions
  • Funders – hence government and society
G8UK - Endorses OA Open Data Charter Policy Paper 18 June 2013

G8UK - Billigt offenen Zugang Eine offene Daten Charter Strategiepapier.
Funder requirements

- UK

  [EPSRC](http://www.epsrc.ac.uk/about/standards/researchdata/Pages/policyframework.aspx)

- USA – NSF, NEH, NIH

- Europe

  [HORIZON 2020](https://ec.europa.eu/research/horizon2020)

  - Most place burden on researcher – some on the institution
RCUK policy - The 1-minute version

• Research data are a public good – make openly available in timely & responsible way
• Have policies & plans. Data with long-term value should be preserved & usable
• Metadata for discovery & reuse. Link publications & data
• Sometimes law, ethics get in the way. We understand.
• Limited embargos OK. Recognition is important – always cite data sources
• OK to use public money to do this. Do it efficiently.
EPSRC policy points

- Awareness of regulatory environment
- Data access statement
- Policies and processes
- Data storage
- Structured metadata descriptions
- DOIs for data
- Securely preserved for a minimum of 10 years from last use

Compliance expected by 2016
Policy and Legal

Policy resources

Overview of Funders' Data Policies
A table and short summaries comparing research funders' policies

Funders' Data Policies
Detailed overview of each funder's policy, stating requirement for data plans, expectations on data sharing and available support.

Institutional Data Policies
A table listing example of UK universities research data policies. Add your examples!

Policy Tools and Guidance
Annotated bibliography of: 1) tools and guidance for creating policies; 2) example policies; 3) publications; & 4) data management guidance.

Preservation policy template
Template to help repositories define preservation policies

Data management plans & DMP Online
Summary of what funders ask for in plans and the DCC's tool to help

http://www.dcc.ac.uk/resources/policy-and-legal
Data reuse stories

• The palaeontologist who saved years of work with archaeological data
• The ‘noise’ from research radar that mapped dust from Eyjafjallajökull
• The 19th-century ships logs that help us model climate change

Often your data tells stories that your publications do not
Compliance

Benefits
Some library roles

• Leadership – coordinate action
• Audit – who has what, where does it go?
• Advice on access – data, wherever it is
• Preservation – permanence
• Citability
• Data/publication linking
• Promoting data in teaching
• Selection
• Education – early career researchers
Who (in the UK) is addressing RDM?
How?

• Create policy – collaborate with others
• Develop existing digital services
• Learn about audit tools (DCC & others)
• Learn about data & sources
• Reskill subject librarians
• Learn about your own data
• Bridge between publishers & researchers
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.
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)
The Data Deluge is upon us

Sensor’s ability to produce data outstrips IT’s ability to process it
Findable, citable data has value

- Important to link publications to data (and vice versa)
- Increases citations – of data & publication
- Increases reuse (hence value)
- But effects exist even without publication, if data is:
  - Archived
  - Citable
  - Discoverable
- All benefit – researcher; institution; publisher

MORAL: build a data registry
How?

• Create policy – collaborate with others
• Develop existing digital services
• Learn about audit tools (DCC & others)
• Learn about data & sources
• **Reskill subject librarians**
• Learn about your own data
• Bridge between publishers & researchers
RLUK/Mary Auckland: Reskilling for Research
9 areas are skill gaps for subject librarians

Sheila Corrall: Libraries, Librarians and Data
Many action exemplars
Other helpful guides

How to Develop Research Data Management Services - a guide for HEIs

Sarah Jones, Graham Pryor and Angus Whyte

CASE STUDY
A Digital Curation Centre Case Study
March 2013

RDM Training for Librarians
Marieko Guy, Digital Curation Centre

Introduction

This case study looks at the approaches taken by three Jisc Managing Research Data Projects (2011 - 2013) and one institution to providing effective training for librarians and information services professionals in Research Data Management (RDM). The case study outlines initiatives and training at the University of Sheffield Library.

Background context

Through its institutional engagement programme the Digital Curation Centre (DCC) has seen many institutional (RDM) initiatives emerge from libraries. Librarians are carving out a new role for themselves in promoting and embedding good RDM practices. They are well placed for this role having information science skills in areas such as metadata, digital preservation, and the repository use - key constituents for RDM. Librarians also tend to have good working relationships with other service departments and researchers. This results in requirements for them to sit on institutional working parties and steering groups redesigning institutional and infrastructure necessary to meet operational and regulatory requirements. Despite these factors various reports indicate that researchers do not always support the library, even when librarians must have a lead role.

The University of Sheffield (UoS) is a member of the White Rose University Consortium's libraries at the Universities of Leeds, Sheffield and York. Development of content and teaching was driven by a need to support the development of the University's research agenda. The role of the library is to provide a service which is relevant to the research needs of the staff and students. The RDM project is led by the University of Sheffield Library with the support of the Jisc Managing Research Data Projects (2011 - 2013) and a range of partners including the University of Sheffield Library and the Library and Information Services Centre.
How?

- Create policy – collaborate with others
- Develop existing digital services
- Learn about audit tools (DCC & others)
- Learn about data & sources
- Reskill subject librarians
- Learn about your own data
  - Help promote data literacy
- Bridge between publishers & researchers
Incremental

The Incremental project is a collaboration between Cambridge University Library and HATII at the University of Glasgow. The project is about engaging with researchers to understand their concerns and needs with regards to data management, namely - how should you create data to make sure in can be found, accessed, understood and reused in the long-term - and

Incremental project

All UK HEIs must take steps to improve data curation activities. The Incremental project is working with researchers at the Universities of Cambridge and Glasgow to build their knowledge of data management, identify their requirements for support and fill any gaps. Findings will be shared to help support other HEIs.

Read more
“Departments don’t have guidelines or norms for personal back-up and researcher procedure, knowledge and diligence varies tremendously. Many have experienced moderate to catastrophic data loss”

Incremental Project Report, June 2010
Excuses – and responses

• “People will ask questions”
  – So use a data centre or repository
• “It will be misinterpreted”
  – Stuff happens. Also, openness encourages correction
• “It’s not interesting”
  – Let others be the judge – your noise is my signal
• “I might get another paper out of it”
  – Up to a point. We might get more research out of it
• “I don’t have permission”
  – A real problem. But solvable at senior level
• “It’s too bad/complicated” – see above
• “It’s not a priority”
  – Unfortunately, funders are making it so. But if you looked at the evidence, it would be your priority as well

See e.g. Carly Strasser’s blog: http://datapub.cdlib.org/2013/04/24/closed-data-excuses-excuses/
What about collaboration?

• Collaborate within the university
• Collaborate with partners
• Collaborate with regional, national services
• Not everything can be done well locally
• Some examples...
Choice of RDM training materials for librarians

2014-04-11

Kevin Ashley – Fiesole2014 - CC-BY
Store your valuable data. Show it to the world. Share it with others.

http://datacentrum.3tu.nl
Data is variable

• Not always textual
• Not always tabular
• Not always fixed
• Not always clearly authored – think of archival provenance
• Not always associated with publication
• There’s more to the world than scholarly research
• Hidden data is wasted data
Closing thoughts

• Library roles: selecting content, protecting it, enabling and encouraging reuse
• All apply equally when the content is data
• Helping users find the most relevant content – much research data does not come from research
• Description can be very different where data is concerned
“Institutions will try to preserve the problem(s) to which they are the solution”

Clay Shirky
Questions

• How does data management align with institutional mission?
• When is library a coordinator, and when is it a service provider?
• What will you do alone, and what will you coordinate with others?
• What skills must you acquire?
• What do you want from national level?
Some conundrums

- Releasing genome data is OK when it’s:
  - An identified human subject
  - An anonymous human subject
  - Your pet dog
  - Another mammal
  - An insect
  - A plant
  - A virus
notes

• No collections – show some data
• Not librarian etc
• Case for data – general, tamiflu
• ‘phd students don’t know what to do’
• Discovery, preservation, permanent ids, selection
• What does it cost/who pays
• Not a publication
• Research data as treasure
• Imperial example of similar time series
• Govt/academic data – the data librarian
• Libraries as providers of data
• Making better use of their own data