Librarians doing data – a paradox?

Graham Pryor
Associate Director, DCC

Libraries@Cambridge, 6th January 2011
The Doomsday Scenario

In 1979 Dennis Lewis, head of ASLIB, wrote what came to be known as the **Doomsday Scenario for librarians**

- This was a time when only basic online information services were beginning to take off
- Lewis said that information professionals wouldn't be around by the year 2000
- Meaning that the *types* of information professionals he saw working in 1979 (mainly librarians and information scientists) would be long gone
- In 1979, with ubiquitous IT services still some way over the horizon, there was a sense that the information age could instead belong to computing scientists, who in turn might enable users to do it all for themselves
What is a librarian?

Cardinal Tommaso Inghirami, Librarian to Pope Leo X – Raphael, 1515-16, Galleria Palatina (Palazzo Pitti), Florence
What is a librarian?

Cardinal Tommaso Inghirami, Librarian to Pope Leo X – Raphael, 1515-16, Galleria Palatina (Palazzo Pitti), Florence

Jeff Haywood, Vice Principal Knowledge Management, Chief Information Officer & Librarian, University of Edinburgh
What is a librarian?

For centuries the custodian of documented knowledge, the librarian has traditionally been the recognised exponent of skills in

- classifying and organising,
- appraising, selecting, annotating,
- preserving, storing and retrieving,
- distributing, sharing and
- managing access to information
The DCC Curation Lifecycle Model

1. CONCEPTUALISE
2. CREATE OR RECEIVE
3. CURATE
4. PREPARATION PLANNING
5. DESCRIPTION
6. REPRESENTATION INFORMATION
7. COMMUNITY WATCH & PARTICIPATION
8. PRESERVE
9. PRESERVATION ACTION
10. INGEST
11. APPRAISE & SELECT
12. REAPPRAISE
13. Migrate
14. Store
15. Access, Use & Reuse
16. Transform
17. Disposal
What is data?

First of all – what is **information** but data that has been contextualised in order to impart a message and what is **knowledge** but what we understand about a subject from the available information

And **data**?
What is data?

First of all – what is information but data that has been contextualised in order to impart a message and what is knowledge but what we understand about a subject from the available information.

And data?

- The lowest level of abstraction in the field of knowledge
- Collections of numbers, characters, images or other symbols that when contextualised in a certain way represent facts, figures or ideas
- Yet in the digital age information and knowledge can be communicated to another person only after they are encoded as data
So what is research data?

- Data that is generated within the context of systematic investigation
- The collected and recorded primary output from a process of observation, experiment or the testing of a hypothesis
- The principal output from scholarly research that when assembled in context and interpreted expertly will produce new knowledge
As an aside - the ‘principal output’? 

Scholarly publications – the tip of visible research output

Research data – the hidden asset (just quietly melting away?)
So…

• If data plus context produces information
• If information properly recorded, represented and understood enables knowledge
• If librarians are the skilled custodians and purveyors of knowledge
• Why, apparently, don’t librarians do data?

• Bearing in mind, of course, that data (hence information) is, in the 21st century, mainly digitally produced and recorded

Is that the problem?
A black hole?

Computing scientists don’t ‘do data’ either.

- They provide and maintain infrastructure for data
- They enable data storage
- They sustain transport mechanisms for data
- They devise software for the manipulation of data

But they don’t ‘do data’ – i.e. they don’t

- Organise, appraise, preserve or curate data

So who does – the data producers?
Incremental Project Scoping Study of Researcher Practice, July 2010

- Creation and organisation of data
  - Inconsistent file structures, naming conventions, etc.
  - Researchers cannot find even their own data
- Data storage and access
  - Many critical research data stored at risk on cheap media with minimal documentation and off network
- Data back-up
  - Data sticks, laptops et al, with few guidelines used
- Preservation of research data – keep it all but…
  - Uncertainty about formats, media and protocols
- Data sharing and re-use – good concept, but…
  - No time for preparation, annotation, contextualising
Researchers and librarians

Patterns of information use and exchange, DCC/RIN, 2009:

• For researchers, data curation is a minor element of the research lifecycle
• Compliance with data management plans is not policed
• The traditional role of professional information intermediaries has been largely replaced by direct access to online resources
• Many researchers have removed themselves from the mainstream library user population
• Substantial discipline knowledge is required of data curators

• Where discipline knowledge is supplied, retention of skills/knowledge is not (Helen Parkinson on data scientists at EBI)
Edinburgh University Data Library

- FINDING…
  “I need to analyse some data for a project but all I can find are published tables and graphs, not the original data source”

- ACCESSING…
  “I’ve found the data I need but I’m not sure how to gain access to it”

- USING…
  “I’ve got the data I need but I’m having trouble analysing it in my chosen software!”

- MANAGING…
  “I’ve collected my own data and I’d like to document and preserve it and make it available to others”

Robin Rice, Data Librarian
Role of libraries

• Training researchers to be more data-aware
• Adopting a data care role via repositories (DISC-UK DataShare project)
• Anticipating increased level of data-related interactional learning and activity between library and research communities
• Data librarians need to be trained on top of discipline background but have no career path yet

Key Perspectives Ltd

http://www.jisc.ac.uk/publications/publications/dataskillscareersfinalreport.aspx
Librarians after Doomsday?

1. Library and Information Science ‘Content specialists’
2. Discipline-based information and knowledge specialists e.g. Data Scientists Subject Librarians
3. Context-specific technology and media specialists e.g. Learning Technologists

Information Technology and Media ‘Conduit specialists’

E-content and digital library specialists e.g. Repository Managers

Hybrid information workers
The challenge for librarians

Patterns of information use and exchange, DCC/RIN, 2009:

• Be proactive! Researchers are usually reluctant to adopt new tools and services unless they know a colleague who can recommend them

• Re-engage with the research community in a way that adds to the effectiveness of research

• Provide specialist support to facilitate the use of new tools, including professional advice, training and documentation on a discipline basis

• Don’t prescribe but recognise the researcher’s strong desire to be consulted about the re-use of data
Incremental Project action plan

Meet researchers’ demands for simple, clear, engaging, and available guidance and support by producing

- Simple, accessible, visual guidance on creating, storing, and managing data with
  - web pages that point to existing local and external resources, illustrated fact sheets, flow diagrams, checklists, and FAQs with solutions to common researcher concerns
- Offer practical data training with discipline-specific examples and assisted by local champions
- Connect researchers with support staff who can offer one-to-one advice and guidance from the proposal writing stage
- Facilitate the development of institutional data management by active communication, awareness raising and building of connections between service and infrastructure providers, policy makers and the research community
So what is a librarian?

Yes, ok, as we’ve seen, someone with skills in

• classifying and organising,
• appraising, selecting, annotating,
• preserving, storing and retrieving,
• distributing, sharing and
• managing access to information

But …
So what is a librarian?

- In a 2009 study of academic librarians, participants were asked to rate the importance of their skills and knowledge on a 5-point scale.
- The five highest ranked choices were:
  - communication and interpersonal skills (4.60)
  - project management/leadership skills (4.56)
  - understanding of digital library architecture and software (4.52)
  - knowledge of the needs of users (4.42)
  - knowledge of technical and quality standards (4.33)

Information Science Today, October 2009

(A traditional blend of skills, moderately revised for 2010)
How to Re-Tool Librarians for Data Curation

Andrew Gallano, Librarian for Digital Services, akg@virginia.edu
Sherry Lake, Scientific Data Consultant, slake@virginia.edu
Scientific Data Consulting Group, Charles L. Brown Science and Engineering Library, University of Virginia, Charlottesville, Virginia, U.S.A.

Abstract

The University of Virginia Library has begun a process of “re-tooling” subject librarians to meet emerging demands of scientific data management. In leading this new initiative, the Scientific Data Consulting Group (SciDaC) focuses on three main activities: 1) “Data Curation Brown Bag” discussions, and 2) data interviews. Through this process, we hope to gradually help traditional subject librarians develop the necessary skills and experience to address data management issues. Each brown bag session focuses on a specific topic (i.e. the NSF data management plan policy, the NSF Data Net Program, etc.) and offers a short presentation and white paper, and then concludes with an informal discussion.

These sessions are expected to help subject librarians become conversant in the issues and to promote the discussion among their departments and faculty. In parallel, the sessions also prepare the subject librarians as partners for the data interviews. Our interview strategy is based on the recent interview models from JISC, Purdue/IN, Oregon and others, with local customizations. We are performing three interviews per month, and each interview consists of two subject librarians, one subject librarian, and the researcher. After several months of leading brown bag discussions and conducting data interviews, we have found that this process is very effective in helping our subject librarians “re-tool” for the new responsibilities. We hope that sharing this model will benefit other institutions as they encounter similar training challenges.

Background on Need for Academic Librarians to Evolve

The library of the future will need subject librarians who are skilled in the area of data curation. Faced with shrinking budgets and staff, institutions may attempt to repurpose existing staff to support researchers and their data services needs (Salo, 2010). Anna Gold’s 2007 paper on infrastructure, data, and libraries, demonstrates that there is a change in academic libraries, with greater emphasis in support of science and engineering scholarship. Garri- tano and Carlson (2005) offer additional examples of how subject librarians should support researchers and their research data. They suggest that librarians need to go beyond introductory articles and reports on data curation to actively collaborate with researchers. They list five sets of skills that librarians need to engage in this new area: 1) library and information science expertise, 2) subject expertise, 3) an ability to build partnerships and develop outreach efforts, 4) a willingness to participate in sponsored research, and 5) an ability to balance workload.

To demonstrate that libraries can provide the needed data curation solutions for researchers, libraries will need to create credible and valuable data services using subject services and data curators (Gardiner, 2009). At the University of Virginia, we have focused on preparing subject librarians for the research data conversation. We have built upon our existing services and developed new skills, roles and organizational structure. Our model does not include adding the full suite of data services to the subject librarian’s list of responsibilities, but instead developing a level of data literacy and then approaching the problem as a team of experts.

Brown Bag Discussions: Every 2 weeks, building subject librarian data literacy

University of Virginia Library Model

In the spring of 2010, the Library decided it was time to go in a different direction and focus more on data management support. This move involved the closing of the Research Computing Lab, and development of a new strategy and plan for research data support. The new vision and strategy, which is now available to the public on the Library website, builds upon the lifecycle framework and now puts more emphasis on the development of institutional repository services and coordinating policy development.

Roles in the Model

Scientific Data Consultants: Work with subject librarians to assess maturity of researcher’s data management practices, make recommendations for how to improve, and shepherd process improvement. Necessary skills: data management, metadata production, data organization/preservation, and systems analysis/design.

Subject Librarians: Serve as subject matter experts and provide opportunities for the scientific data consultants to work with researchers and graduate students. Necessary skills: collaborative, ability to build partnerships, communication, content expertise.

Current Training Model and Support

1) “Data Curation Brown Bag” Discussions: The SciDaC group hosts “Data Curation Brown Bag” sessions every two weeks. The objective of these brown bag lunches is to:
   a. Gradually educate the subject librarians on the pressing issues and trends facing data curation.
   b. Provide an opportunity for a short presentation and white paper (1 page, very straightforward overview of the issue), and then casual discussion.
   c. Help the subject librarians become comfortable with data curation issues, and then promote the discussion among their departments and faculty as they interact with them.

2) “Data Interviews” Goal: is to develop an understanding of how our science and engineering researchers manage their research data and initiate a discussion about how to simplify processes and improve practices. Each interview is scheduled for 45 minutes and includes the Scientific Data Consultants, the Subject Librarian and the researcher. A final interview report is then distributed to all Subject Librarians helping to give them a better understanding of research data processes beyond their own field:
   a. Identify common research data problems and needs.
   b. Identify the types of digital “data” that is being created.
   c. Identify communities and individuals who are at the most pressing issues.
   d. Develop opportunities to provide data management recommendations and training.

Goals and Objectives

1) Build data literacy for subject librarians
2) Develop knowledge of how researchers actually manage their data at U.Va.
3) Create opportunities to consult and collaborate
4) Establish the Library as the place for preservation of data

References Cited

Retooling at the University of Virginia

Faced with shrinking budgets and staffs, institutions may attempt to repurpose existing staff to support researchers and their data services needs (Salo, 2010).

Roles in the Model

- **Scientific Data Consultants**: Work with subject librarians to assess maturity of researcher’s data management practices, make recommendations for how to improve, and shepherd process improvement. Necessary skills: data management, metadata production, data organization/preservation, and systems analysis/design.

- **Subject Librarians**: Serve as subject matter experts and provide opportunities for the scientific data consultants to work with researchers and graduate students. Necessary skills: collaborative, ability to build partnerships, communication, content expertise.
Creating a Data Management Plan & Other Training Opportunities in the Library

Introduction

Researchers in large academic institutions have multiple barriers to identifying available campus tools and resources for managing their research data. The recently announced National Science Foundation (NSF) data management plan requirement for all new grant applications is a prime opportunity to address these difficulties. The University of Minnesota Libraries has been developing educational and outreach programs to support researchers throughout the data life-cycle.

User Needs

Our 2006 and 2007 studies of researchers across campus identified unmet needs for assistance with data management and opportunities for the libraries to take a lead role. To quote one College of Biological Sciences faculty member commenting on data organization:

There are probably better ways. If there were a workshop on organization and file management, I would go. The Libraries do this so well. 

A 2008 survey of nearly 800 researchers on campus explored these needs in greater depth. There was clear evidence of an education gap in the way researchers manage their data. For example, over a quarter had lost important data due to the lack of backup and nearly half used insecure, external hard drives for backing up data (see Figure 1). 

![Figure 1. Data backup methods used by University of Minnesota faculty](image)

Our Approach

Our response has been multi-pronged and recognizes the need to work with partners across campus.

Face-to-face workshops and consultations

- Introduction to Data Management
  - Practical data management problems
  - Campus and other data storage options

Creating a Data Management Plan for Grant Applications

- Overview of questions to address
- One-on-one data sharing consultation
- Plan to work with Office for Vice Provost of Research to include in PI continuing education requirements

Online resources and tools

![Figure 2. Managing Your Data website screenshot](image)

Future Directions

Web-based tool guides researchers through data management plan creation.

- Provides directory of data-related campus services
- Poses questions based on NSF requirements

Lessons Learned

1. Based on the evaluation of our workshop participants we have adapted our approach to include more tools specific to managing data, such as TortoiseSVN version control for software, and demonstrations of file format conversion to preservation-friendly formats.

2. Promotion is key. In order to reach audiences, we partnered with the graduate school to promote our workshop.

3. Users prefer online workshop for introductory skills. We saw a drop in attendance once recording was posted online.

4. Knowledge of discipline specific resources are key. We are building librarian subject expertise in these areas.

References


Lisa Johnston, Physics & Astronomy Librarian (ljohnsto@umn.edu)
Meghan Lafferty, Chemistry Librarian (mlaffert@umn.edu)
There are probably better ways [of managing data]. If there were a workshop on organization and file management, I would go. The Libraries do this so well — Faculty Staff, College of Biological Sciences

- **Face-to-face workshops and consultations**
  - Introduction to Data Management
  - Practical Data Management Problems
  - Campus and other Data Storage Options
  - Creating a Data Management Plan for Grant Applications
  - One-on-one data sharing consultation

- **Online resources and tools**
Managing Your Data

Digital data is growing at an exponential rate: from the digital family photos on a home computer to the terabytes of data generated by researchers in the various disciplines across the university. How do we as individuals and scholars in the e-research environment keep up with our growing data management needs?

The University Libraries are working toward sustainable digital stewardship through best practices of research data management and awareness of open data issues. This site examines the life-cycle of researcher data and offers tools and solutions for creation, storage, access, analysis, and dissemination.

On this page:
* Why Data Management?
* Creating a data management plan
* How can the library help me manage data?
* Beyond research data: E-scholarship

Contact Us
- Lisa Johnston, Physics, Astronomy and Geology Librarian
- Meghan Lafferty, Chemistry, Chemical Engineering and Material Sciences Librarian
- Amy West, Social Sciences and Data Services Librarian
Help is at hand!

Prioritisation of resources
Capacity development and planning
Efficiency savings – move data to more cost-effective storage
Manage risks associated with data loss
Realise value through improved access & re-use

Scale: research teams, departments, institutions
DCC Events

Research Data Management Forum
Targeted at data managers, practitioners

DCC Roadshows
Effecting institutional change in data management

International Conference
Forum for research in data curation
DCC Helpdesk

- info@dcc.ac.uk
  - For answers to questions about all aspects of our work
- http://www.dcc.ac.uk/
  - For news, resources, tools and much more