The Research Data Lifecycle and Data Management

Possible Roles and Opportunities for Librarians and Publishers

Presented by ASME as a Service to Our Librarian Community

ASME
SETTING THE STANDARD

Contents

"Interest in research data has grown substantially over the past decade. The reason for this is evident: the digital revolution has made it far easier to store, share, and reuse data." [1]

-- Joyce M. Ray, PhD, MLS
   Program Coordinator, Digital Curation Program
   Johns Hopkins University

"Data needs to be collected, documented, described, and made accessible in ways that extend its use beyond the original research focus, so it can be employed in ways even the researcher didn't envision." [2]

-- Amy Nurnberger, MSc
   Program Head for Data Management Services
   MIT Libraries

[1] Ray, Joyce M., ed., 2014, *Research Data Management: Practical Strategies for Information Professionals, Purdue University Press, West Lafayette, Indiana.*

[2] "Q&A with Amy Nurnberger" in the *Bibliotech Newsletter,* MIT Libraries, last modified October 17, 2017, accessed December 19, 2018, **https:// libraries.mit.edu/news/bibliotech/with-nurnberger**

Introduction

The era of Big Data has arrived.
And with it are new possibilities to work alongside the research community.

It is ASME's intention that this document will serve as a catalyst for discussion among librarians, researchers, and publishers. It is noted that currently the adoption of RDM practices can vary by geographic region and type of institution.

Researchers in the academic, government, and corporate sectors are generating massive quantities of data across all scientific, technical, and medical (STM) disciplines at an accelerating rate. Increasingly, government and other funding bodies require data management plans so that the results of funded research are publicly available.

This environment has created both opportunities and challenges for **librarians and information specialists**, who support **researchers and the research process** by collecting, managing, and sharing this data, and for **publishers**, who are often responsible for disseminating this data so that they are easily accessible, executable, and reproducible.

Along with this data explosion comes the growing demand for intelligent and effective Research Data Management (RDM).

RDM is the planning, organization, sharing, and preservation of data, from when it enters the research lifecycle through to the dissemination and archiving of results. It is essential to ensure that data can be preserved and remain accessible (and reproducible) in the long term so that beyond the completion of a research project or expiration of its funding, the resulting data can serve as the foundation for verification or future research.

Librarians are being asked to play an important role to help manage the data essential for driving innovation forward — and the possibilities within this role are expanding!

Librarians and information specialists have the opportunity to develop and implement services offered to your research community within an evolving scholarly landscape and facilitate access to and preservation of scholarly communication.

What Is Big Data?

The buzzword BIG DATA is firmly entrenched in our understanding of the Information Age. It refers to extremely large and complex datasets, whether structured, semi-structured, or unstructured, that may be analyzed computationally to reveal important patterns, trends, and associations.

Big Data Can Be Characterized by the "4 Vs"
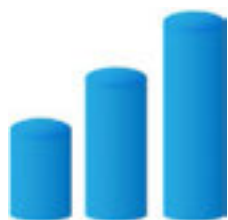
Volume
Scale of data

Velocity
Speed of data

Variety
Diversity of data

Veracity
Certainty of data

Big Data and Research Data Management

Good Research Data Management (RDM) practices are designed to help researchers, librarians – and publishers – overcome the challenges created by Big Data, and at the same time, realize the benefits and insights that it provides in the pursuit of scientific, technical, and medical advancements.

RDM continues to evolve to address a wide range of information needs. To some extent, it is influencing the practices of researchers, librarians, and publishers to think about the end result in advance.

RDM is typically driven by these 7 pillars at an institutional level: [3]

1. **Storage:** The need to provide immediate storage facilities for a wide variety of datasets at a scale that anticipates the future requirements of researchers, ensures against data loss with adequate back-ups, and in a way that represents value for money and is convenient to use.

2. **Security:** The requirement to ensure that data (particularly confidential or sensitive) should be held securely with relevant authentication and authorization mechanisms in place.

3. **Preservation:** The need for medium and long-term archiving of data with associated selection protocols and preservation activities along with a supporting technical infrastructure.

4. **Compliance:** The need to comply with the requirements and policies of other relevant agencies, particularly funders, as well as legal obligations, such as data protection, and industry good practice.

5. **Quality:** The imperative to maintain and enhance the quality of research activity in general in order to demonstrate the robustness of findings and enable results verification and reproducibility (partly derived from, but not limited to the quality of research data itself).

6. **Sharing:** The need to share data among targeted users and also to provide mechanisms and systems to enable open access to data where appropriate.

7. **Jurisdiction:** The development of a professional narrative around the need to be involved in RDM and how this impacts other stakeholders in the institution.

The Research Data Lifecycle

The key to successful RDM begins with understanding the research data lifecycle and how librarians and information scientists can contribute to this process.

The essential steps in the research data lifecycle are well documented, and there are numerous models. Depicted below is the UK Data Service model (**www.ukdataservice. ac.uk/manage-data/lifecycle**), which serves as the basis of this discussion.[4]

---
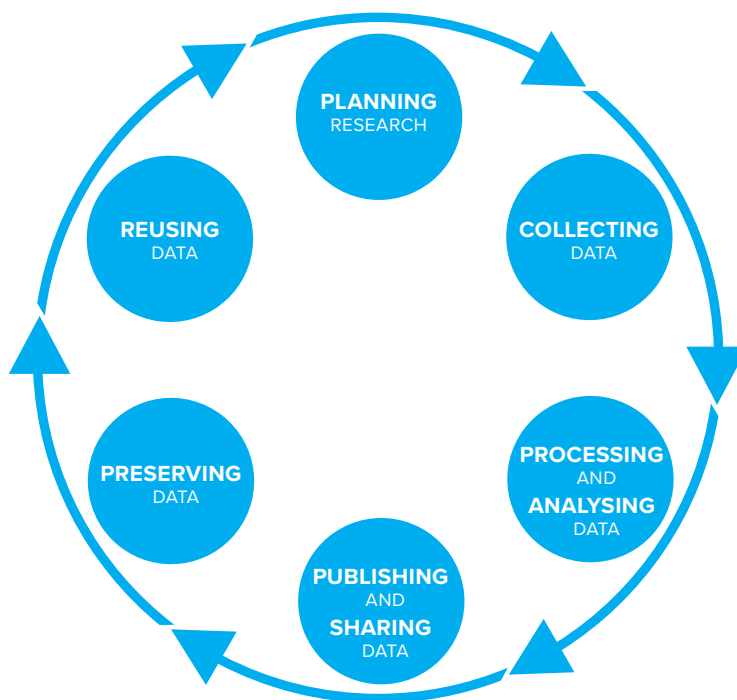
[3] Based on: Pinfield S., Cox, A.M., and Smith, J., 2014, "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences," PLoS ONE 9(12). DOI: 10.1371/journal.pone.0114734.

[4] "UK Data Service: Research Data Lifecycle," accessed December 17, 2018, **https://www.ukdataservice.ac.uk/ manage-data/lifecycle**

Librarians and Publishers: Opportunities and Shared Roles

Contributing to the research data lifecycle in different ways, from different perspectives, librarians and publishers are uniquely positioned to support researchers at every stage by providing more sophisticated services, tools, and analytics to manage data. Dialog and collaboration among these groups (with librarians as advocates for certain practices) are growing and are essential to help overcome the challenges inherent to the explosion of research data.

> Thinking beyond purely traditional roles, how can librarians and publishers help sustain and enrich the research process? Here are some possibilities that highlight new opportunities and expanded roles for librarians and publishers. Some of these tasks could be shared or interchanged between librarians and publishers.

Planning Research

Recommended Steps:

- Design research
- Plan data management
- Plan consent for sharing
- Plan data collecting, processing, protocols, and templates and associated expenses
- Explore existing data sources

What Librarians Can Do:

- Assist in writing grant applications
- Participate in data management plan consultation
- Create specific recommendations for research groups concerning data management training/consultation

> **TIP FOR LIBRARIANS:**
>
> Get involved in the planning/ proposal writing stage as early as you can so you can lend your expertise to the development of the data management plan.

How Publishers Could Contribute:

- Provide guidelines for data reproducibility including availability, metadata, preservation, citation, etc.
- Provide guidelines for how data should be made available
- Provide trusted (perhaps neutral) places to host datasets

Collecting Data

Recommended Steps:

- Collect data
- Capture data with metadata
- Acquire existing third-party data

What Librarians Can Do:

- Provide guidelines or advice on documentation/metadata schema
- Determine data needs, including finding and possibly licensing third-party data

> **TIP FOR LIBRARIANS:**
>
> Begin thinking about how to manage and license both internal and external data and metadata and the associated costs and security issues, keeping the needs of the collection and researcher in mind.

How Publishers Could Contribute:

- Provide guidelines or advice on documentation/metadata schema
- Provide linking between published articles and underlying data
- Allow text mining via API

Processing and Analyzing Data

Recommended Steps:

- Enter, digitize, transcribe, and translate data
- Check, validate, clean, anonymize
- Derive data
- Describe and document data
- Manage and store data
- Analyze and interpret data
- Produce research outputs
- Cite data sources

What Librarians Can Do:

- Recommend tools and resources so data can be easily organized and shared and fulfill funding and other compliance requirements
- Advise on use of relevant tools (by assisting in search) and highlight the problem of reinvention of code
- Ensure that research data outputs are discoverable and accessible
- Advise on ethical and copyright issues and the area of personal data
- Ensure results are properly curated

> **TIP FOR LIBRARIANS:**
>
> If data analysis is outside of your skill-set, look for tools or develop computational and statistical expertise by furthering your professional education.

How Publishers Could Contribute:

- Provide tools to assist with analysis, including alternative metrics
- Ensure results are publishable and discoverable
- Assist with technical, domain-specific "know how"

Librarians and Publishers: Opportunities and Shared Roles

Publishing and Sharing Data

Recommended Steps:

- Establish copyright
- Create metadata to facilitate discovery
- Select appropriate access to data
- Publish/share data
- Promote data

What Librarians Can Do:

- Verify institutional and funder requirements or restrictions
- Help create user documentation
- Support open science
- Help researchers share their results and/or prepare for journal submission

> **TIP FOR LIBRARIANS:**
>
> **Help researchers understand the publishing ecosystem, particularly in today's evolving open access environment.**

How Publishers Could Contribute:

- Serve as the basis of published scholarly communication, primarily in the form of journal articles, that can allow for publication of underlying datasets
- Disseminate quality, citable research globally
- Support open science

Preserving Data

Recommended Steps:

- Migrate data to best format/media (including versioning checks)
- Store and back up data
- Create preservation documentation
- Preserve and curate data

What Librarians Can Do:

- Define what is in scope at an institutional level for data preservation
- Advise on preservation strategies for both born-digital and digitized material
- Integrate preservation and cataloging paradigms due to knowledge and experience of preservation standards
- Train researchers in data management and workflows and advise on facilitation of workflows
- Train in use of software curation tools and making data discoverable
- Provide institutional repositories and assistance to properly utilize them

> **TIP FOR LIBRARIANS:**
>
> **The value of data should be analyzed to determine suitability for long-term curation and preservation in accordance with documented policies, guidance, or funding requirements.**

How Publishers Could Contribute:

- Ensure non-published outputs of research are accessible and future-proofed
- Provide a framework for discussions between publishers and librarians about depositing academic research into institutional archives designed to help address common workflow issues
- Ensure that end products are citable (this is an incentive to curate data)

Reusing Data

Recommended Steps:

- Conduct secondary analysis
- Undertake follow-up research
- Conduct research reviews
- Scrutinize findings
- Use data for teaching and learning

What Librarians Can Do:

- Encourage researchers to think about how their data can be reused at the start of their research process
- Maintain an infrastructure that supports the preservation and retrieval of datasets
- Discuss the barriers that the academic promotion and tenure process creates for those who want to freely share their data
- Help researchers understand and plan for the lifecycle of data

> **TIP FOR LIBRARIANS:**
>
> **It is important to encourage researchers to better manage their data so others can reuse it — to advance the discovery to the benefit of all people.**

How Publishers Could Contribute:

- Make it easy for researchers to meet funder requirements
- Encourage the peer review of data sets when it is practical to do so
- Become proactive "data stewards"

Looking Towards the Future

- **Where is RDM heading?**
- **What technological advancements could potentially change the playing field?**
- **How will AI impact RDM?**
- **What role will open science play?**

These are only a few of the uncertainties facing all stakeholders concerned with RDM. But what is certain is the role of and opportunities afforded to librarians continue to evolve and expand in new and exciting ways. The expertise, skillsets, and talent for problem-solving that librarians already possess can be leveraged to support effective RDM now and in the future. Marketing a library's additional services is a good way to increase awareness of the contributions its staff can make to planning and implementing effective RDM.

Librarians, researchers, and publishers will need to increasingly engage in collaborative dialog that fosters strong partnerships in order to help achieve their institutions' or companies' goals.

> **TIP FOR LIBRARIANS:**
>
> Librarians can use their current experience in new ways to develop new skills and position themselves for career advancement.

References, Resources, and Tools

References

[1] Ray, Joyce M., ed., 2014, *Research Data Management: Practical Strategies for Information Professionals*, Purdue University Press, West Lafayette, Indiana.

[2] "Q&A with Amy Nurnberger" in the *Bibliotech Newsletter*, MIT Libraries, last modified October 17, 2017, accessed December 19, 2018, **https://libraries.mit.edu/news/bibliotech/with-nurnberger**

[3] Pinfield S., Cox, A.M., and Smith, J., 2014, "Research Data Management and Libraries: Relationships, Activities, Drivers and Influences," PLoS ONE 9(12). DOI: 10.1371/journal.pone.0114734.

[4] "UK Data Service: Research Data Lifecycle," accessed December 17, 2018, **https://www.ukdataservice.ac.uk/manage-data/lifecycle**

Resources

ASME Data Policy: Data Archiving and Sharing, **https://www.asme.org/shop/journals/information-for-authors/data-policy**

Corti, L., Van den Eynden, V., Bishop, L., and Wollard, M., 2014, *Managing and Sharing Research Data: A Guide to Good Practice*, Sage Publications Ltd, London.

"Glossary of Data Management Terms," Cornell University Research Data Management Service Group, **https://data.research.cornell.edu/content/glossary**

"Data and Reproducibility, Committee on Publication Ethics: COPE." **https://publicationethics.org/data**

Krier, L. and Strasser, C.A., 2014, *Data Management for Libraries: A LITA Guide*, ALA Tech Source, Chicago.

Piwowar, H. A. and Vision, T.J., 2013, "Data Reuse and the Open Data Citation Advantage," PeerJ, e175, **https://doi.org/10.7717/peerj.175**

Springer, R. and Cooper, D., March 14, 2019, "Guest Post — Pictures Worth a Thousand Words? On Visualizations of Scholarly Workflow," *The Scholarly Kitchen*, **https://scholarlykitchen.sspnet.org/2019/03/14/guest-post-pictures-worth-a-thousand-words-on-visualizations-of-scholarly-workflow**

Tools

DCC (Digital Curation Center) – Provides expert advice and practical help to anyone in UK higher education and research wanting to store, manage, protect, and share digital research data, **www.dcc.ac.uk**

DMP Tool (Data Management Plan Tool) – Guides researchers on how to create, review, and share data management plans that meet institutional and funder requirements, **https://dmp.cdlib.org**

ESIP Federation – A collection of training videos about very specific topics from "Tracking Data Usage" to "Handling Sensitive Data," **http://commons.esipfed.org/datamanagementshortcourse**

DataOne Data Management Modules – 10 PowerPoints accompanied by handouts and hands-on exercises, **https://www.dataone.org/education-modules**