RESEARCH DATA MANAGEMENT

A BRIEF OVERVIEW FOR PHYSICAL SCIENCES LIBRARIANS

PRESENTED BY AIP PUBLISHING AS A SERVICE TO LIBRARIANS

Scientists at academic and government institutions and corporate laboratories are generating data at unprecedented levels, at an increasingly accelerated pace, and with consistently decreasing costs. This explosion of born-digital research data (data that are created in digital form) means that the era of BIG DATA has arrived. Along with this digital overload comes the growing need for intelligent and effective Research Data Management (RDM).

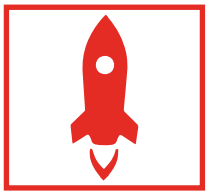
WHY IS THIS IMPORTANT FOR LIBRARIANS?

BIG DATA is creating new roles and opportunities for physical science librarians and information professionals across all subject disciplines working alongside the research community.

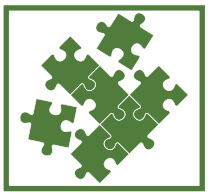
WHAT IS BIG DATA?

BIG DATA refers to extremely large data sets, whether structured or unstructured, that may be analyzed to reveal important patterns, trends, and associations.

BIG DATA CAN BE CHARACTERIZED BY THE 4 Vs



VELOCITY
Speed at which data is generated



VARIETY
Compiling different types of data that all work together

VOLUME
Massive amounts of data

VERACITY
Trustworthiness and quality of data

WHAT IS RDM?

Research Data Management is defined as the organization and description of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. [Whyte A, Tedds J (2011) ***Making the case for research data management***. http://www.dcc.ac.uk/webfm_send/487]

RDM addresses a wide range of information needs and is typically driven by:

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 and in a way that represents value for money and is convenient to use.
2. **Security:** The requirement to ensure that data, particularly that which is 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 amongst 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 upon other stakeholders in the institution.

[From Pinfield S, Cox AM, Smith J (2014) ***Research Data Management and Libraries: Relationships, Activities, Drivers and Influences***. PLoS ONE 9(12): e114734. DOI: 10.1371/journal.pone.0114734]

A NEW ROLE FOR LIBRARIANS AND INFORMATION PROFESSIONALS? THE INFORMATIONIST

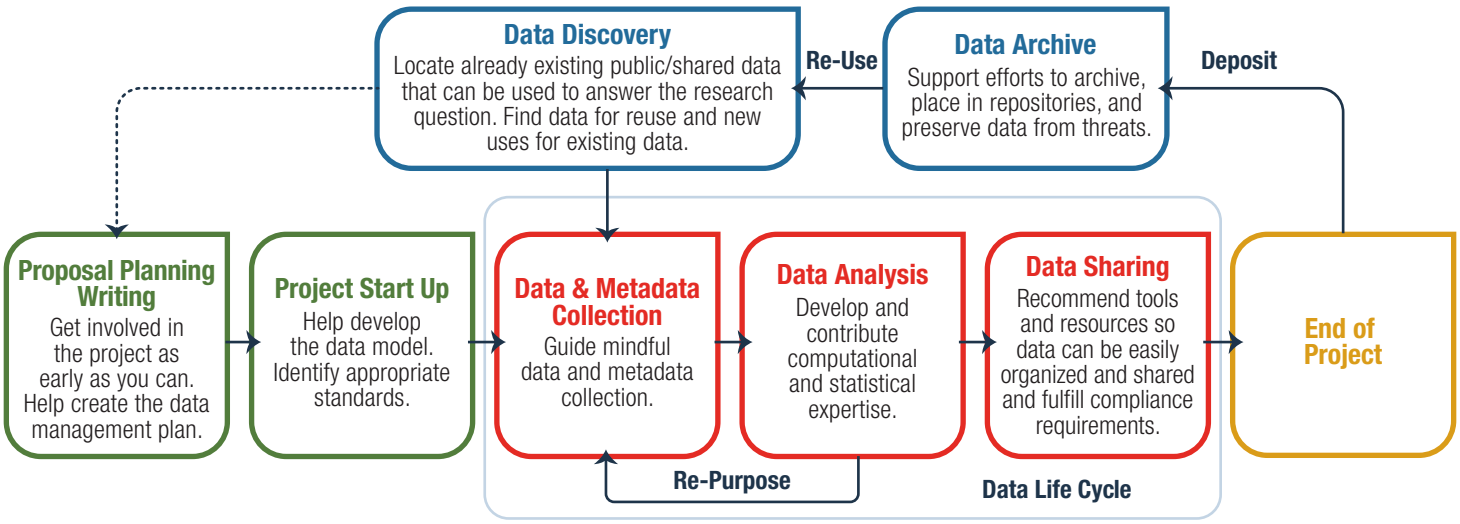
Researchers can be overwhelmed by the sheer volume of data and how to organize, analyze, and share it in meaningful ways, which in many cases is mandated by institutional and government funding agencies. **Librarians and information professionals can help. For those of you who would like to take on this challenge, your training, experience, and talent for problem solving have prepared you to take on the proactive role of "informationist."**

Whether an RDM program is initiated by the library or results from a request from the research community, a librarian with subject discipline knowledge is typically embedded within a research group, working hands-on with its data. This brings the "power of the librarian" right into the scientists' environment and leverages library expertise.

Skills that physical sciences librarians already possess can be leveraged to support effective RDM

HOW CAN LIBRARIANS CONTRIBUTE TO RDM?

Here are just some of the ways in which librarians can make significant contributions to the RDM process:



Adapted from: University of Virginia Library - http://data.library.virginia.edu/data-management/lifecycle

WHAT DATA MANAGEMENT TOOLS CAN HELP SUPPORT RDM?

Two valuable resources that focus on helping researchers understand how to effectively manage their data are:

- **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

Marketing your library's nontraditional services is a good way to make your institution aware of the contributions you can make to effective RDM

WHAT DOES THE FUTURE HOLD FOR LIBRARIES AND RDM?

There's no doubt that BIG DATA is a *big deal* and will continue to grow. Perhaps your physical sciences research community is already benefiting from RDM programs in which librarians are actively involved. It's a win-win: Libraries contribute directly to driving research forward and receive recognition from the administration. **Librarians use current expertise in new ways, develop new skills, and position themselves for career advancement.**

WANT TO FIND OUT MORE?

Pinfield S, Cox AM, Smith J (2014) ***Research Data Management and Libraries: Relationships, Activities, Drivers and Influences.*** PLoS ONE 9(12). DOI: 10.1371/journal.pone.0114734

Kratz JE, Strasser C (2015) ***Researcher Perspectives on Publication and Peer Review of Data.*** PLoS ONE 10(2). DOI: 10.1371/journal.pone.0117619

AIP Publishing extends its thanks to Lisa Federer, MLIS, MA, AHIP, Research Data Informationist, Office of Research Services, National Institutes of Health, Bethesda, MD, USA, for sharing her insights on RDM.