20th International Conference on Grey Literature - December, 3-4 2018
Loyola University, New Orleans, Louisiana USA

The data librarian: myth, reality or utopia?

Silvia Giannini, Anna Molino
CNR-ISTI, Italy

Outline

1. Scenario and objectives
2. What is research data?
3. The Research Data Management (RDM)
4. Digital curation
5. Which professional figures?
6. Case studies
7. The *data librarian* profile
8. Conclusions

Scenario

- Open Science
  - Open Access to scientific publication.
  - Open Data as a *primary research output.*
- Requests for services and infrastructure to manage and preserve research data from many funding bodies .
- Academic libraries as "*aggregator, collector and curator of external scholarship, be it printed or online*".

Objectives

- Who and how is managing the research data?
- What experiences worldwide?
- What roles for academic libraries?

What is research data?

- The definition and the management of research data are depending on the differences between disciplinary field and organizational context
  - *observations, images, computer program results, recordings, measurements, experiences, theory, test or hypothesis, or another research output is based…*
- Research data comes in a variety of formats
  - *numerical, descriptive, visual or tactile…*
- It may be raw, cleaned or processed.
- It may be held in any format or media.

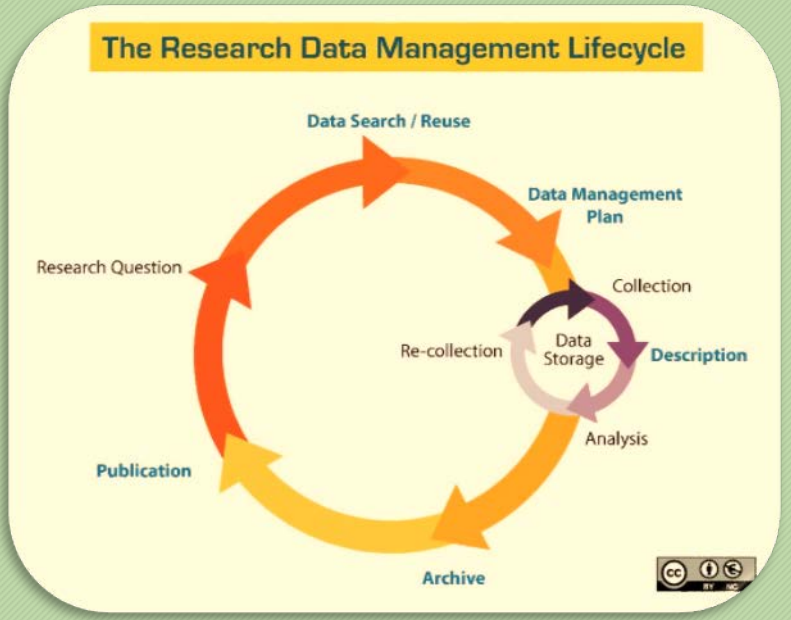
Research data is the material underpinning a research assertion

The Research Data Management (RDM)

<<Research data management concerns the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.>> Whyte A., Tedds J. (2011)

RDM as Research Data Service (RDS)

- Data Management Plan
- Digital curation
  - *selection*
  - *preservation*
  - *maintenance*
  - *archiving*
- Metadata creation and transformation



Research Data Management Lifecycle, diagram The University of California

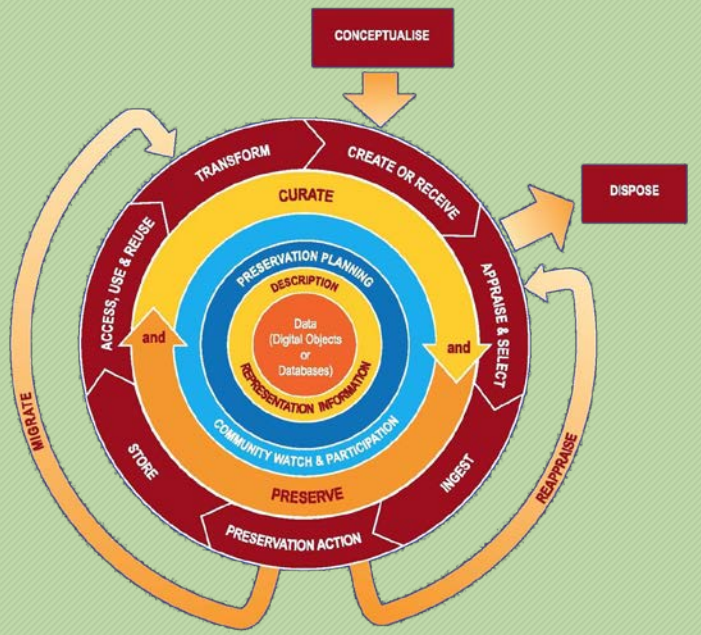
Digital curation

- The concept considers the ubiquity of digital contents and their need to be preserved, managed and valued.
- *Digital curation* is
  - a broader concept;
  - a *cross-activity* among different domains;
  - involving the entire *lifecycle of a digital resource.*

- *The DCC model summarizes the full lifecycle actions in:*
  - description and representation information;
  - preservation planning;
  - community watch and participation;
  - curate and preserve.

The need to identify a professional figure who manage and store the growing amount of data in digital format has generated the role of ***digital curator***.

Digital Curation Centre (DCC) Curation Lifecycle Model



www.dcc.ac.uk

Which professional figures?

The *digital curators* community

The link of the *digital curation* with the development and management of Open Access repositories makes *data curation* a subset of the *digital curation.*

At the same time, we can consider *data curator* a specialization of the *digital curation* figure, while...

Thus, data curation is

- a narrower concept;
- strongly connected to the academic world;
- referred to data as *research data*;
- involving the entire *lifecycle of digital research data.*

...the wide experience of librarians in different disciplinary domains, their skills in the management of metadata sets and in the selection, care and maintenance of collections directly refers to the figure of the ***data librarian***.

Case study – University of Edinburgh Research Data Service (RDS)

RDS: tools, support and training to aid staff and students working with research data

RDM support and consultancy

- Before: Data Management Plan.
- During: data search, storage, sharing, and versioning.
- After: datasets recording; open data sharing; data archiving.

Training

- Online:
  - MANTRA[1] free online course for those who manage digital data.
  - Special focus "DIY Training Kit for Librarians[2]."
  - RDMS MOOC [3](Massive Open Online Course) in collaboration with the University of North Carolina.
- Face to Face: workshops and courses.

[1]https://mantra.edina.ac.uk/

[2]https://mantra.edina.ac.uk/libtraining.html

[3]http://www.coursera.org/learn/data-management

Case study – The RDNL FO-BO Model

Research Data Netherlands (RDNL): alliance of DANS, 3TU.Datacentrum and SURFsara.

Services provided: data curation, management and storage.

Front Office (FO)

- ❑ Data management: information provision and researchers training.
- ❑ Temporary storage facilities and data transfer.
- ❑ Data acquisition.

Back Office (BO)

- ❑ Data management: data stewardship, long-term storage and accessibility.
- ❑ Training and support to FO data librarians.
- ❑ Secure sustainable storage and retrieval at the end of the research project.

- ❑ Specific FO-BO responsibilities vary from organization to organization.
- ❑ Duties of front and back offices may overlap: coordination and role definition are necessary.
- ❑ Institutes performing FO tasks only may outsource their BO tasks to a data archive.

Case study – eResearch at Griffith University

Division of Information Services (INS): library and Information Technology (IT) services.

Librarians: acquisitions, collection development, copyright advice, IT training, OA, publication repositories, research assessment exercises, bibliometrics.

eResearch Services: build and manage technical infrastructures for supporting researchers.

Librarians: "generalists" as they need to have a wide range of skills, knowledge and expertise

Core set of skills
- ❑ Advanced metadata skills
- ❑ High level communication skills
- ❑ High level documentation skills

Core set of knowledge
- ❑ Knowledge of the broader researcher environment
- ❑ Knowledge of the research process and of scholarly communication
- ❑ Knowledge of legal and regulatory framework

Generic skills and knowledge
- ❑ Technical skills (e.g. system design, business analysis, usability testing)
- ❑ Project Management to understand project lifecycles, PM methodologies, internal PM frameworks and reporting requirements

The *data librarian* profile

<<"Traditional" librarian competences and skills into renewed organizational structures.>>

(*Schmidt et al., 2016*)

Core competences (Schmidt & Shearer, 2016)

- ❑ Provide access to data.
- ❑ Advocacy and support for managing data.
- ❑ Managing data collections.
- ❑ Related services: OA and institutional repositories; collection development; advisory services (e.g. copyright policies); information literacy; digital curation and preservation; digital collections.

eResearch role (Brown et al., 2015)

- ❑ Understanding of discipline-based research process, outputs and scholarly communication (e.g. data types and formats).
- ❑ Knowledge of ethics, intellectual property, copyright and licensing.
- ❑ Knowledge of discipline-specific metadata schemas and related standards (item and collection level).
- ❑ Knowledge of repository certification schemes and standards.
- ❑ Knowledge of semantic web standards.

Conclusions

Echo and Narcissus by John William Waterhouse

The Radical Christianity of Thomas More's "Utopia"

- The importance and need for data management is recognised and established by national entities and funding bodies.
- It arises from the case studies that having a formal RDM Policy leads to the definition of specific roles involved in RDS.
- While roles and professional skills of the *digital librarians* can be often considered as re-interpretations of traditional roles, the role of the *data librarian* is one of the most complex and should be defined completely.
- The worldwide scenario shows data librarians operating in various contexts with different backgrounds.

So, *neither myth… nor utopia …but a composite reality*