Workshop

A Curriculum for Research Data Management:
how do you educate the Data Scientist?

Seamus Ross – University of Toronto

Donatella Castelli – ISTI CNR

Vittore Casarosa – ISTI CNR

Anna Maria Tammaro – Universiy of Parma

Conveners

13:30-13:45 Introduction

13:45-15:00 Invited speakers
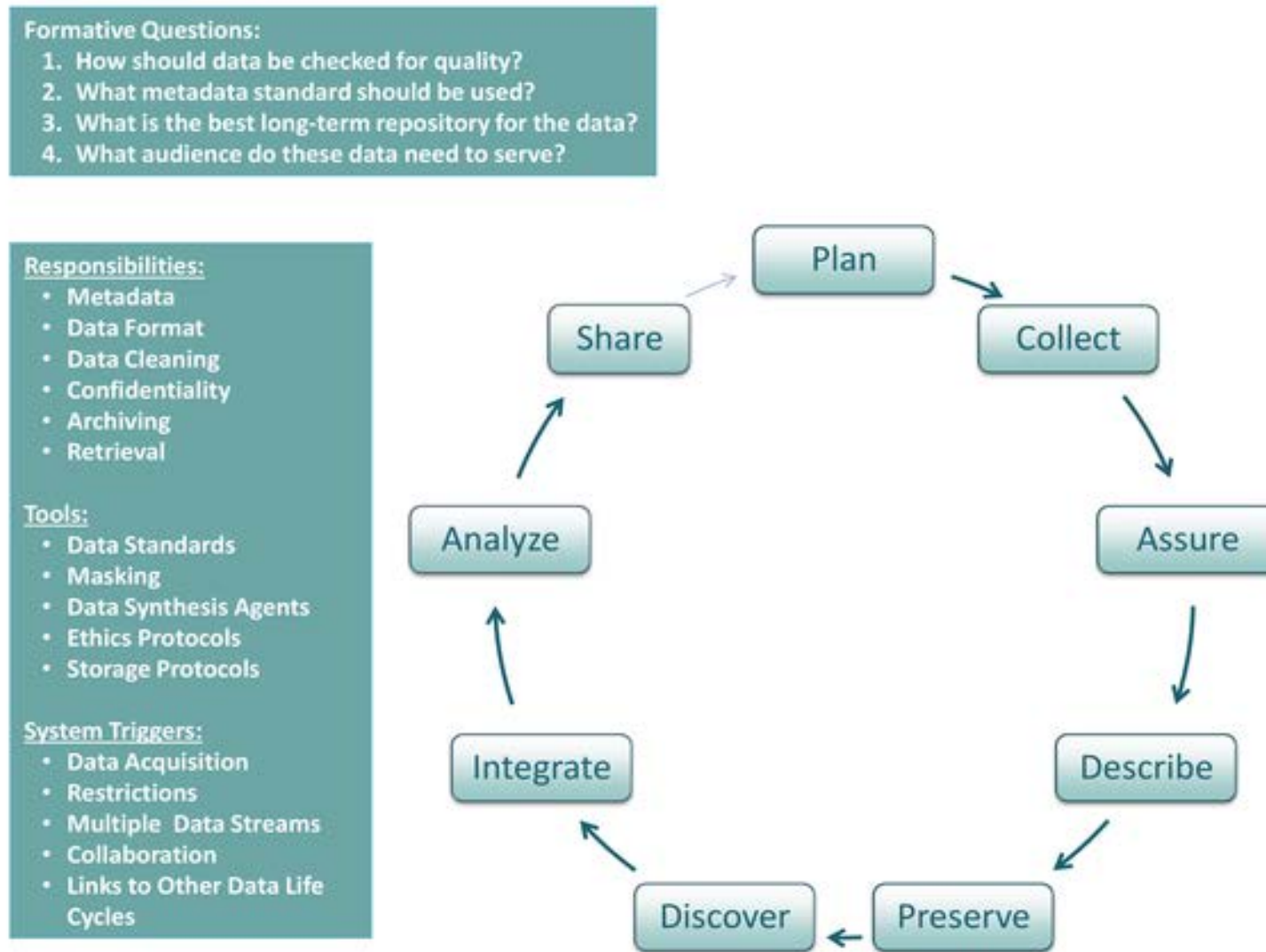
15:00-15:30 Break

15:30-16:30 Wold Cafe discussion

16:30-17:00 Summary discussion and conclusion

Programme

Open Science

It is expected that the publication and dissemination of research results will include, together with the 'traditional' paper, the data used to reach the results presented in the paper and possibly also a description of the experiments (if any) that have led to those results.

Figure 1. The life cycle of data: the steps needed to responsibly collect, record, store, and steward data.

Formative Questions:
1. How should data be checked for quality?
2. What metadata standard should be used?
3. What is the best long-term repository for the data?
4. What audience do these data need to serve?

Responsibilities:
- Metadata
- Data Format
- Data Cleaning
- Confidentiality
- Archiving
- Retrieval

Tools:
- Data Standards
- Masking
- Data Synthesis Agents
- Ethics Protocols
- Storage Protocols

System Triggers:
- Data Acquisition
- Restrictions
- Multiple Data Streams
- Collaboration
- Links to Other Data Life Cycles

Plan → Collect → Assure → Describe → Preserve → Discover → Integrate → Analyze → Share → Plan

PLOS | BIOLOGY

Data curator: who is he/she?

- In recent times we have seen an increased use of the buzz words digital curation and data curator.

- The exact meaning given to those terms usually depends on the background and the context of the speaker and the listener.

Convergence? New identity?

- How those competencies could play in favour of the convergence of the three traditional "memory disciplines" (libraries, archives, museums), given the same digital nature of the resources to be curated

- Is it a different professional identity, given the different focus and mission of the three disciplines, where the value adding and the access to the collections would remain different?

Definition

The term Data Curator is more and more used to indicate the person/organization responsible for all the activities connected with the management (curation) of research data, which can take many different aspects, such as previous publications, images, video, audio, data bases, email, web sites, etc., and most of the time those data are specific to the research field.

Data scientist

- it is not (yet) clear which of the existing professional roles are best suited for this activity. Should there be a Data Librarian, or a Data Archivist, or a Data Museum curator ? Or should the responsibility of curating research data be given to the data producers, i.e. the researchers themselves ? Or is this a new role to be invented from scratch, with the emerging new name of Data Scientist?

Invited lectures

- **Michael Seadle**

  - Technical skills in Library and Information Science classrooms

- **Wolfram Horstmann**

  - A Data Science preparatory course

- **Gobinda Chowdhury**

  - Data Science: a new discipline or an agglomeration of skills?

"World Cafe'"

- Who is the Data Scientist: a librarian? an archivist? a researcher in the discipline? a computer scientist? Do we really need a Data Scientist?

- What actually does the Data Scientist? How can a Data Scientist be educated and trained?