Research Data and Data Management Planning

Markus Stocker

September 12, 2017

Outline

- What are research data
- Research data lifecycle
- Data types, formats, models and standards
- Metadata
- Data management, plans and planning tools

Datum is ultimately reducible to a lack of uniformity [1]

Define data

- Entities, physical or digital, used as evidence of phenomena [2]
- A reinterpretable representation of information [3]
- ...
- There is no consensus definition
- Even institutions that curate data may not define what they curate

Data examples

- Not just spreadsheets of numbers, also
- Sequences of bits
- Characters on a page
- Recording of sounds
- Physical and biological specimens
- Images
- Software

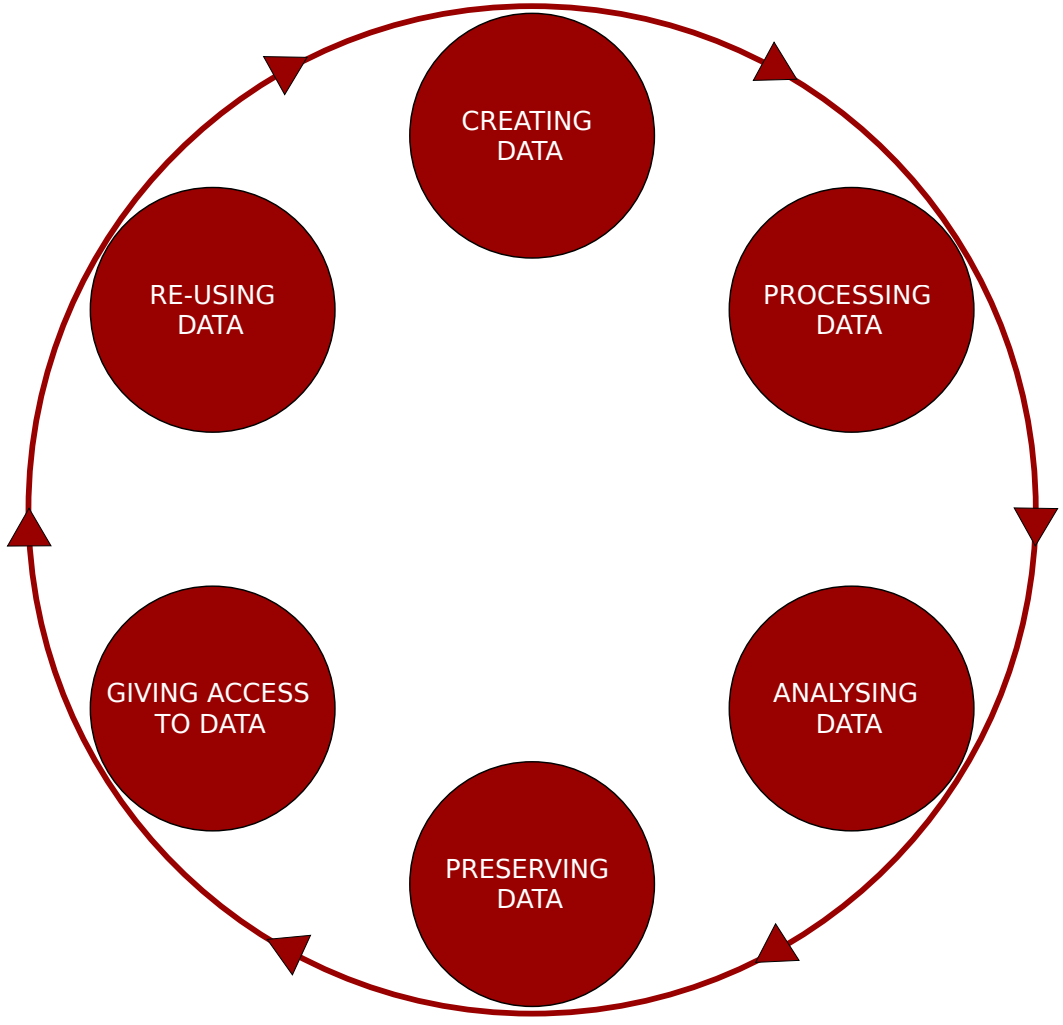
Define research data

- Unsurprisingly, there is no consensus on the definition
- Factual material [...] necessary to validate research findings [4]
- Everything needed to reproduce a given scientific output [5]
- ...

Research data examples

- In addition to the obvious, e.g. data files
- Notebooks, e.g. laboratory, field, diaries, ...
- Questionnaires, audio and video tapes
- Models and scripts
- Workflows and protocols

Research data lifecycle



Adapted from `http://www.data-archive.ac.uk/create-manage/life-cycle`

Research data types

- Observational data
  - Result from recognizing, noting or recording facts
  - Collected by human observation, surveys, instruments
  - Typically difficult or impossible to reproduce
  - Measurement of ocean temperature at locations in space-time
- Experimental data
  - Result of procedures in controlled conditions
  - In theory reproducible but may be expensive
  - Chemical analysis in a laboratory
- Computational data
  - Result in executing computer models, simulations, or workflows
  - Reproducible if software and input available
  - Plant disease pressure model

Data formats

Data models

-

Standards

-

Metadata

-

Data management

-

Planning data management

-

Tools for data management planning

-

Take aways

References

[1] Luciano Floridi. *The Philosophy of Information*. Oxford University Press, 2011. ISBN 978-0-19-923239-0.

[2] Christine L. Borgman. *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press, 2015. ISBN 9780262028561.

[3] CCSDS. Reference Model for an Open Archival Information System (OAIS). Recommended Practice CCSDS 650.0-M-2, The Consultative Committee for Space Data Systems, Washington, DC, USA, June 2012. URL `https://public.ccsds.org/Pubs/650x0m2.pdf`.

[4] EPSRC. Research Data. URL `https://www.epsrc.ac.uk/about/standards/researchdata/scope/`.

[5] Alisa Surkis and Kevin Read. Research data management. *Journal of the Medical Library Association : JMLA*, 103(3): 154–156, jul 2015. doi: 10.3163/1536-5050.103.3.011. URL `https://doi.org/10.3163/1536-5050.103.3.011`.

Slide 3: Joan Miró (1968). Landscape. Acrylic on canvas. Fundación Joan Miró, Barcelona.
`https://www.fmirobcn.org/en/colection/catalog-works/5442/p-landscape-p`