RESEARCH & INNOVATION
excellence, innovation, leadership.

Why manage research data?

Scientific data curation, citation and scholarly publication

Prof Lessing Labuschagne
Executive Director: Research, University of South Africa

Learn without limits.

Research Spending

UK	• £3.5 billion spent on research undertaken by UK universities (2012)
USA	• $55 billion spent on science and engineering alone (2009)
European Commission	• €50 billion (£42.4/$61.5 billion) spent on research (2007 – 2013)
SA	• R22.2bn spent on R&D (2011/2)

Citation

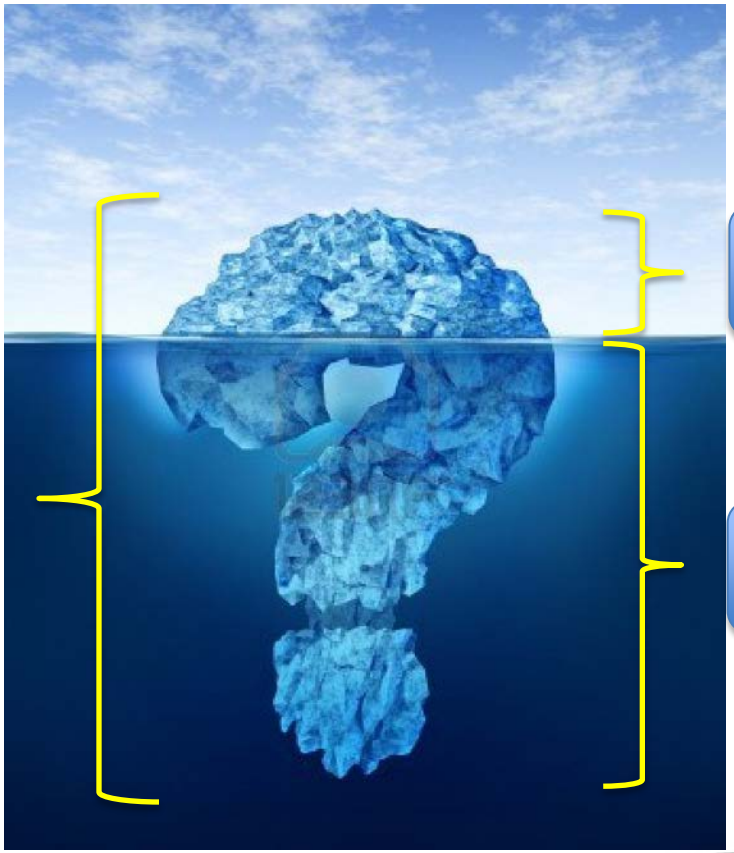
- “Are 90% of academic papers really never cited?”
- “We Must Stop the Avalanche of Low-Quality Research”
  - 60% of social and natural science articles never cited?

- Non-citation rates vary enormously by field
  - 12% of medicine
  - 82% for the humanities
  - 27% for natural sciences
  - 32% for social sciences

Articles as self-contained resources

59,487 researcher

R22,1bn



Research Outputs

12,363 (2012)

Research Data

?

Research Data

- *Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property*

Common Principles on Data Policy Research Councils UK (RCUK, 2011)

Research Lifecycle

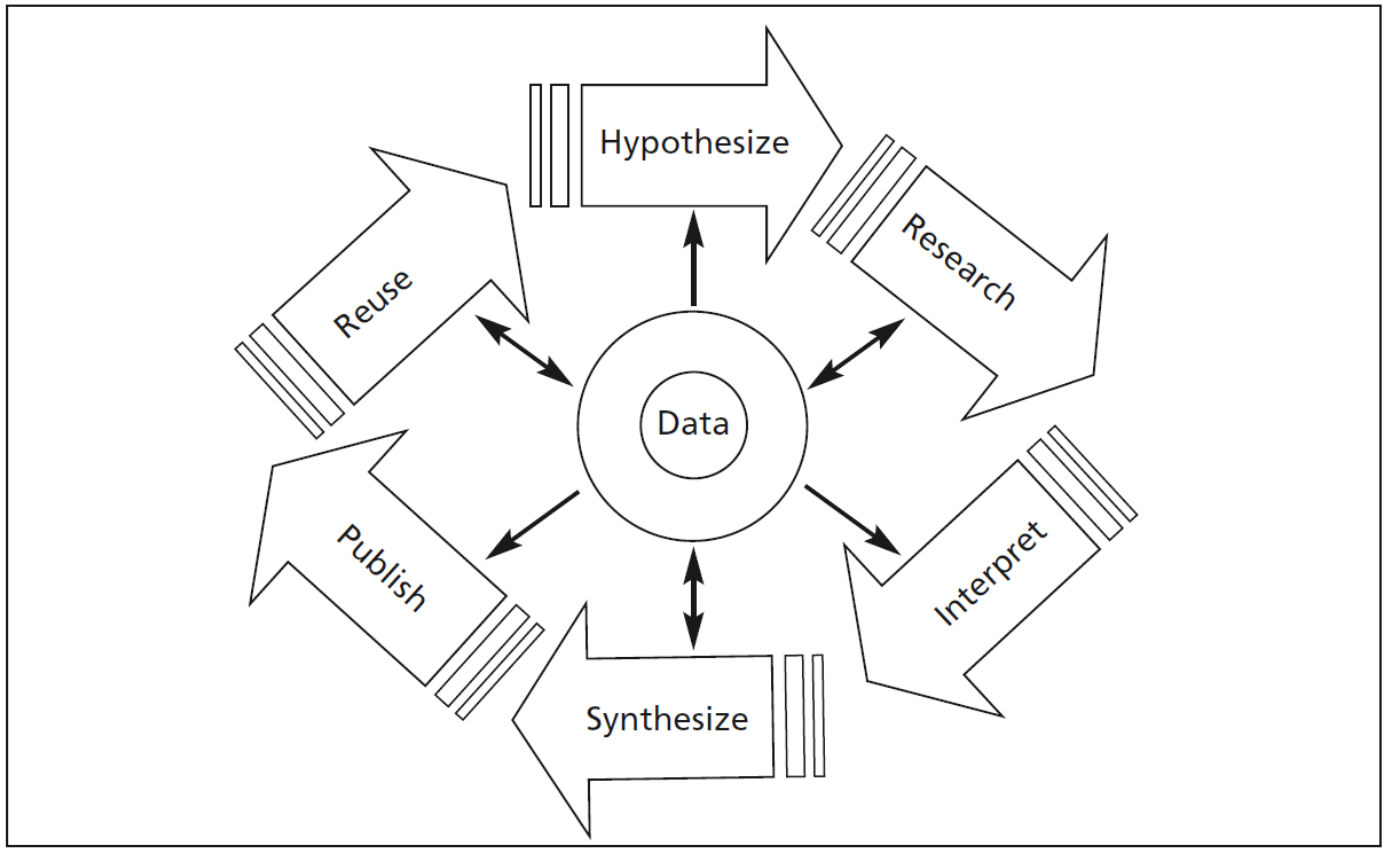


Figure 1.1 *The six datacentric phases of the research lifecycle*

Managing research data, Pryor, G., 2012

Data

- Producing data at increasing orders of magnitude
- Scientific research has become increasingly data-intensive
  - annual rate of increase = 30%
- Biosciences
  - raw image files for a single human genome = 28.8 terabytes, (30,000 gigabytes)
- High energy physics
  - Large Hadron Collider (LHC) experiment at the European Organization for Nuclear Research (CERN), in Geneva produces 15 petabytes (15 million gigabytes) of data annually
  - 1.7 million dual-layer DVDs

Scientific Data

- Once published, scientific data should remain available forever so that other scientists can reproduce the results and do new science with the data
- Data may be used long after the project that gathered it ends
- It is likely that new techniques of data production and manipulation will still be developed

Scholarship

- Shift from a document-centric view of scholarship to a data-centric view of scholarship
- For example
  - Social scientists are accessing data from fields such as the health sciences and environmental sciences and using tools such as geographical information systems to study the connection between health and personal relationships or environmental conditions.
  - New research has been enabled by the digitization of weather records extracted from a previous century of ships' logs, with data not originally gathered for that purpose now being used in research into climate change.

Why manage research data?

- Meet funding body grant requirements
- Ensure research integrity and replication
- Ensure research data and records are accurate, complete, authentic and reliable
- Enhance data security and minimise the risk of data loss
- Prevent duplication of effort by enabling others to use your data

Managing Data

- Account for content and context
- To understand the data, those later users need the metadata:
  - (1) how the instruments were designed and built
  - (2) when, where, and how the data was gathered
  - (3) a careful description of the processing steps that led to the derived data products that are typically used for scientific data analysis.
- It is now feasible and economical to store everything

Research Value Chain

- There are several roles in the research process: *Authors*, *Publishers*, *Curators* and *Consumers*.
- There is pressure on Authors to publish their research in comprehensible ways and there is demand from Consumers for these publications

Evolution of librarians

- Preserving documents that have been the foundation of civilizations
- Traditional custodians of documented knowledge, with their armoury of skills in appraising, classifying, preserving, storing and retrieving information
- Researchers uniformly expressed a need for organizing, describing, managing, archiving, and accessing data

Curation

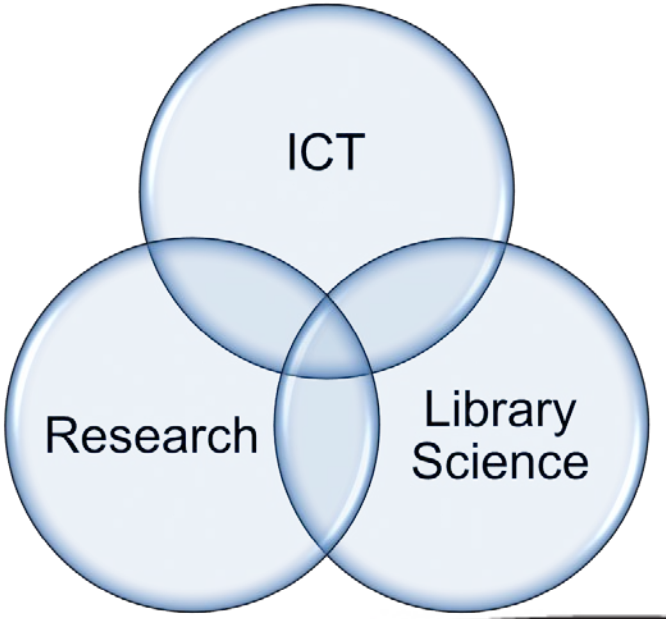
- Data curation - a means to collect, organize, validate, and preserve data so that scientists can find new ways to address the grand research challenges that face society

- Digital preservation – securing the persistence of information in digital form
  - authenticity and integrity
  - trustworthiness of digital resources
  - organisation and long-term archiving
  - added-value services and new uses for the resources
  - knowledge enhancement and presentation

- Future fitness of digital information

Conclusion

- Librarians should embrace the role of data curator to remain relevant and vital to our scholars



Learn without limits.

UNISA university of south africa

Challenges

- Few Standards : There are few guidelines for publishing data. There are fewer metadata standards.
- Laborious: It is laborious to document the data and the data reduction process.
- Confidentiality / Privacy
- Research ethics
- Legislation
- A personal ideology disposed toward sharing
- While digital storage space is not an issue, server maintenance and management are on-going problems.
- proprietary data or funding agencies that require non-disclosure agreements for all or part of the research or data
- Extremely competitive field or disciplinary culture that discourages outside involvement.
- which materials are appropriate for inclusion in the archival store
- Ownership / acquire custody of the items / sufficient intellectual property rights
- Data quality