Data

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

RDM and the role of the information professional

Presentation at NeDICC Workshop on 'The role of the information professional in relation to research data management', CSIR, Pretoria, South Africa

Johann van Wyk

10 June 2016

Introduction

"Researchers are concerned with a number of important research data management issues such as access, organisation, analysis, storage, combination and re-use, portability, sharing, and data security; in addition they frequently struggle to manage the data (often in huge quantities) that they collect" (Auckland, 2012: 21).

This presents an opportunity for Information Professionals to be increasingly involved in supporting researchers to improve the discoverability of both their research data and their publications.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

What is meant by 'Research Data'?

Research data, unlike other types of information, is collected, observed, created or generated, for purposes of analysis to produce original research results

http://www.docs.is.ed.ac.uk/docs/data-library/EUDL_RDM_Handbook.pdf

What is Research Data Management?

- "Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results.

- Research Data Management is part of the research process, and aims to make the research process as efficient as possible, and meet expectations and requirements of the university, research funders, and legislation"

University of Leicester
http://www2.le.ac.uk/services/research-data/rdm/what-is-rdm

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Why manage research data?

- Meet funding body grant requirements, e.g. NSF, NIH;

- Meet publisher requirements

- Ensure research integrity and replication;

- Increase your research efficiency;

- Enhance your reputation as researcher, and increase your citations

- 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;

- Save time and resources in the long run;

- Comply with practices conducted in industry and commerce;

- Protect institutions from reputational, financial and legal risk

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

NRF Statement on Open Access

- "From 01 March 2015, authors of research papers generated from research either fully or partially funded by NRF, when submitting and publishing in academic journals, should deposit their final peer-reviewed manuscripts that have been accepted by the journals, to the administering Institution Repository"

- "the data supporting the publication should be deposited in an accredited Open Access repository, with the provision of a Digital Object Identifier for future citation and referencing".

http://www.nrf.ac.za/media-room/news/statement-open-access-research-publications-national-research-foundation-nrf-funded

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Research Data Lifecycle

- Data Repurposing/Re-use
- Data Citation

Re-using Data

Creating Data

- Designing Data Management Plans
- Data Capture
- Data Storage
- Metadata creation
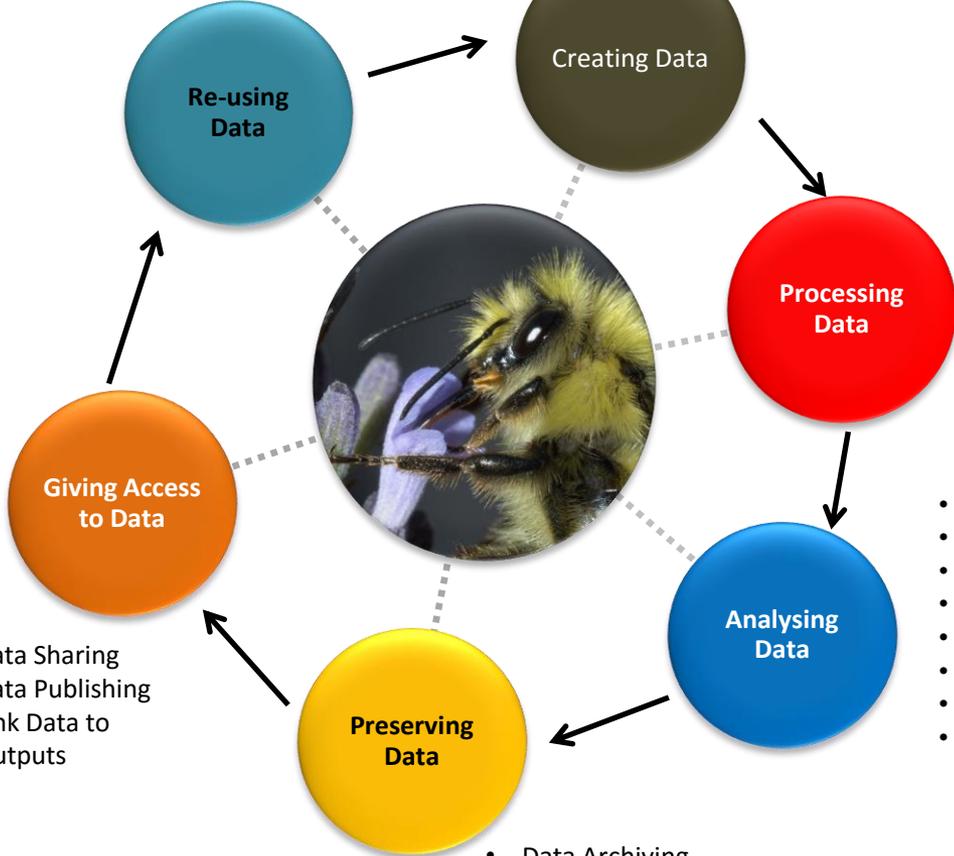
Processing Data

- Data Storage
- Metadata Creation
- Data Cleansing
- Data Verification
- Data Validation
- Data Anonymisation

Giving Access to Data

- Data Sharing
- Data Publishing
- Link Data to Outputs

Preserving Data

Analysing Data

- Data Storage
- Metadata Creation
- Data Cleansing
- Data Verification
- Data Validation
- Data Anonymisation
- Data Interpretation & Analyis
- Data Visualisation

- Data Archiving
- Data Preservation
- Metadata Creation
- Link data to outputs

Based on UK Data Archive Lifecycle

Research Data Lifecycle in the context of the Research Life Cycle
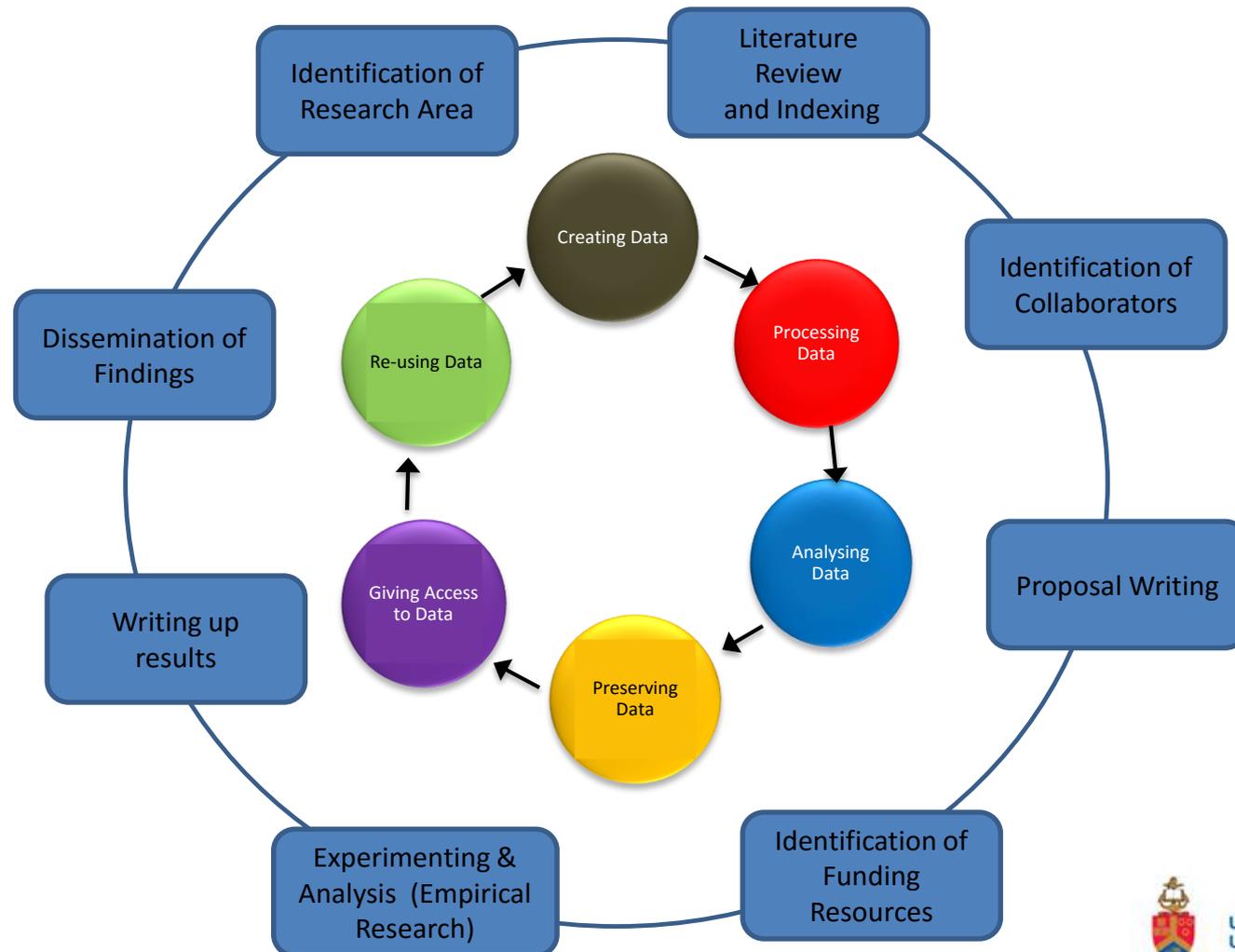
The role of Information Specialist in RDM

General Tasks

- Conduct Data Interviews: See Data Curation Profiles Toolkit (Purdue University) http://datacurationprofiles.org/

- Advise lecturing staff, graduate students, and administrators about scholarly communication issues, e.g. Open Access, the institutional repository, and social tools such as Researchgate, etc.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
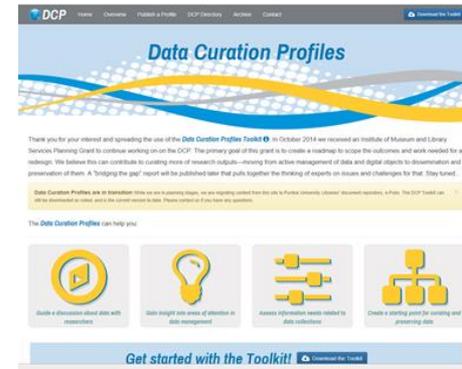
Designing Data Management Plans

- A Data Management Plan is "a formal document that outlines what you will do with your data during and after you complete your research" (The University of Virginia Library, 2014).

- **DMPonline tool (pilot for South Africa)** http://pilot.meraka.csir.co.za

- Information Specialists could play an advisory role
- Information Specialists could train researchers in the DMPonline tool
- Information Specialists could support researchers in complying with the various open access mandates of funding bodies

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Capture/Collection

The action or process of "<u>gathering</u> and <u>measuring</u> information on variables of interest, in an established systematic fashion that enables one <u>to answer</u> stated research questions, test hypotheses, and evaluate outcomes" (Responsible conduct of research, n.d.; The Oxford Dictionary, 2014).

Examples of data collection methods:

Observations, literature study, textual or visual analysis, interviews, surveys, experiments, etc.

Librarians can play their traditional role of information searching, - training and – consultation

Searching for data sets? Go to re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

Or Data Citation Index(Web of Science)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Storage/Back-up

Data storage is the process of "<u>preservation</u> of data files in a secure location which can be <u>accessed readily</u>".

Data Backup is the process of "preserving <u>additional copies</u> of your data in a separate physical location from data files in storage". (Research Data Services, University of Wisconsin-Madison, 2014)

Information Specialists could advise:
- On File Naming Conventions -
  See File Naming Rules, University of Edinburgh
  http://www.ed.ac.uk/records-management/records-management/staff-guidance/electronic-records/naming-conventions
- Data protection (Strategies to keep data safe)
- Organising their data folders and files

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Metadata Creation

Metadata is searchable, standardised and structured "<u>information that describes a dataset</u>" and explains "the aim, origin, time references, geographic location, creating author, access conditions and terms of use of a data set"

(Corti et al., 2014: 38; USGS Data Management Website, 2014)

Examples: Dublin Core; PREMIS

Information specialists, and especially cataloguers have the skill-set to assist with metadata creation and to advise. This will improve the discoverability of researchers' research data and their publications

Sources on Metadata and Cataloguing:

- BOYDSTON, J.M.K. AND LEYSEN, J.M. 2006. Observations on the catalogers' role in Descriptive Metadata Creation in academic libraries. *Cataloging & Classification Quarterly*, Vol. 43, No. 2,p. 3-17.

- MILSTEAD, J. AND FELDMAN, S. 1999. Metadata: cataloging by any other name…*Online*, Jan/Feb 1999, Vol. 23, no. 1, p.24-31

- *Rethinking Technical Services: new frameworks, new skill sets, new tools, new roles.* Edited by Bradford Lee Eden. Lanham: MD: Rowman and Littlefield, 2015.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Types of Metadata

```
Data Retention ─┬─ Descriptive Metadata ── Dissemination Phase
                ├─ Structural Metadata ── Research In Process - Phase
                └─ Administrative Metadata ── Preservation Phase
```

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
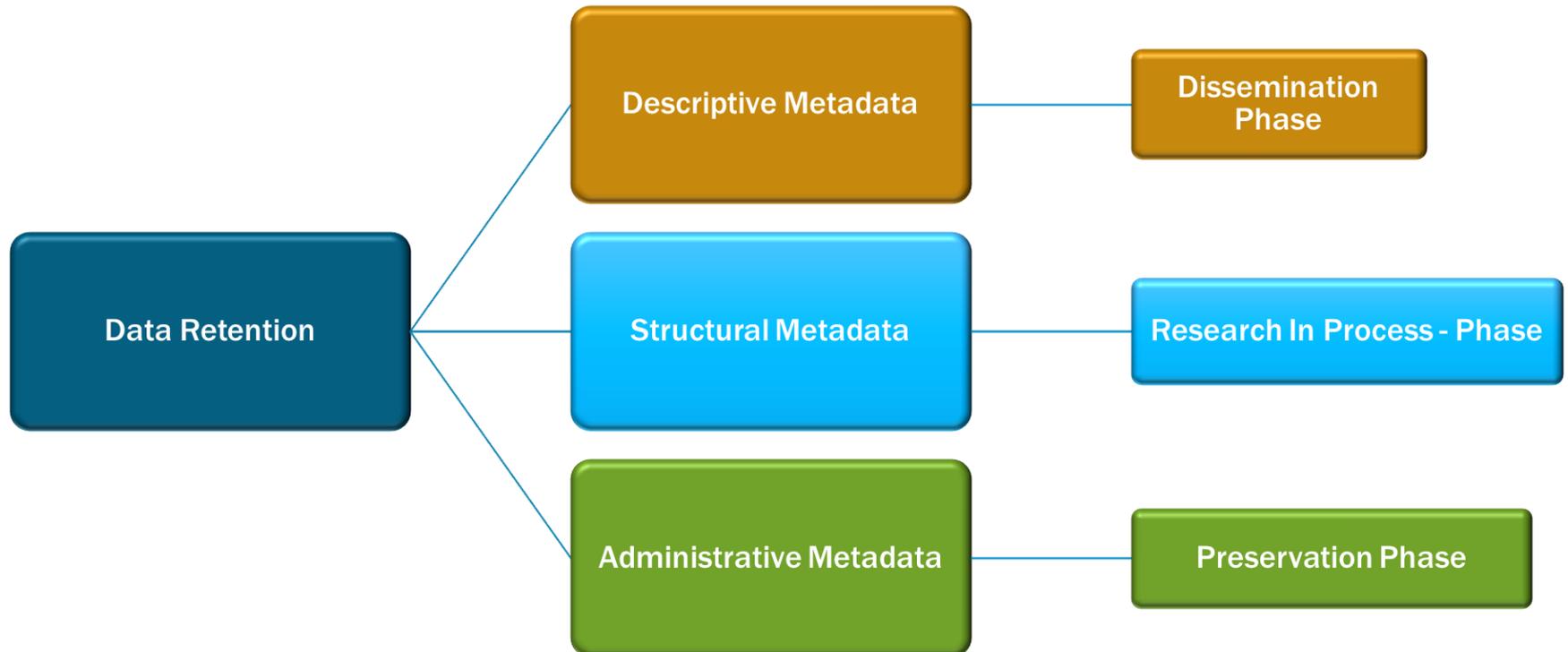
Types of Metadata

Descriptive Metadata

describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.

Structural Metadata

Structural metadata describes the physical and/or logical structure of digital resources

Administrative Metadata

Administrative metadata often captures the context necessary to understand information resources, such as creation or acquisition of the data, rights management, and disposition.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Interpretation and Analysis

Data interpretation and analysis "is the process of assigning meaning" to the gathered information and ascertaining "the conclusions, significance, and implications of the findings" (Analyzing and Interpreting Data, n.d.).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Visualisation

Data Visualisation is the <u>visual representation</u> of data, and is used to enable people to both understand and communicate information through <u>graphical</u> and <u>schematic</u> avenues (Friendly, 2009: 2; Schnell and Shetterley, 2013: 3)
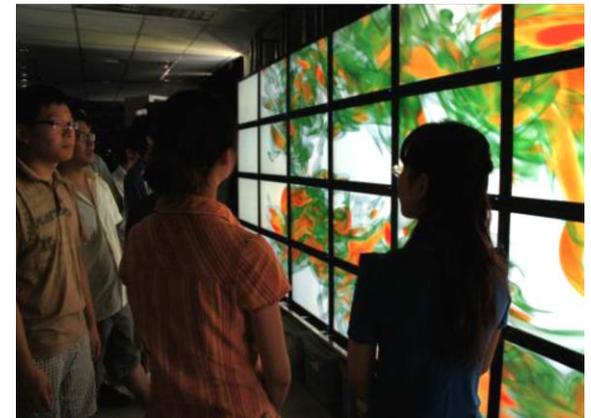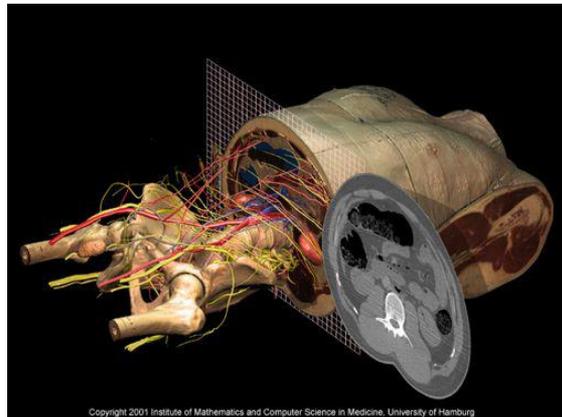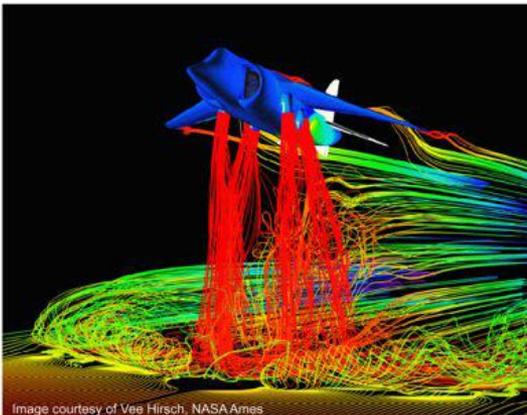
From Xiaoru Yuan's presentation, CODATA Workshop on 12 June 2014

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Processing Data

Analysing Data

- **Data Cleansing**

"refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this dirty data' (Wikipedia)

https://powerofus.force.com/articles/Resource/Focus-on-Your-Data

- **Data Verification**

"the process of evaluating the completeness, correctness, and compliance of a dataset with required procedures to ensure that the data is what it purports to be. (Martin and Ballard, 2010: 8-9; US EPA, 2002:7)

- **Data validation**

process "to determine if data quality goals have been achieved and the reasons for any deviations. Validation checks that the data makes sense". (Martin and Ballard, 2010: 8; US EPA 2002:15).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Publishing (Sharing)

- **Data publishing**

This is the process of making research data underpinning the findings published in peer-reviewed articles, available for readers and reviewers in an appropriate repository, or "as supplementary materials to a journal publication" (Corti et al 2014: 197; Marques, 2013)

- **Data Journals**

A more recent development. These journals publish data papers that describe a dataset, and also give an indication in which repository the dataset is available (Corti et al. 2014: 7-8).

A list of Data Journals – available at
http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Publishing
(Role of the Information specialist)

Information Specialists:

- Could advise on IP rights, data protection, copyright, plagiarism as well as ethical issues around their data
- Could advise and give training on uploading in a data repository
- Could advise and give training on how to cite Data Sets, and how to create identifiers (DOIs)
- Could advise on how to <u>license</u> research data., e.g. Creative Commons etc.
- Could advise on embargoes to data
- Could advise on tools to track impact (bibliometrics): e.g. http://impactstory.org

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Linking Data to Research Outputs

This is the process of <u>connecting</u> the underlying <u>data</u> relating to a specific <u>research output</u>, e.g. journal article, thesis, etc to the research output itself. This can be done by adding a <u>digital object identifier (DOI)</u> to the dataset and including this in the metadata of the research output, or by <u>citing</u> the dataset (Callaghan et al., 2013).

The Librarian could assist researchers, through training and consultation on DOIs and data citation/referencing methods

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
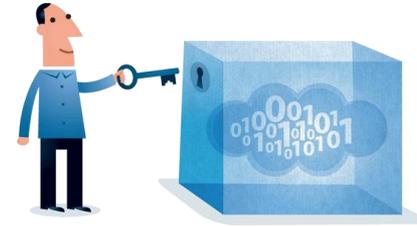
Data Preservation

Image from www.digitalbevaring.dk

- Data preservation is "the process of providing enough <u>representation</u> <u>information</u>, <u>context</u>, <u>metadata</u>, <u>fixity</u>, etc. to the data and [thereby securing the data] so that anyone other than the original data creator can use and interpret the data" (Ruth Duerr, National Snow and Ice Data Center as cited by Choudhury, 2014)

The Information Specialist
- Could assist researchers in preparing data for long-term preservation, by advising on metadata standards
- Could provide advice on where to submit data for preservation
- Could provide advice on preservation file formats

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data Citation

Data citation is the process of <u>referencing</u> (attributing and acknowledging) reused data in a similar fashion as traditional sources of information (Corti et al. 2014: 197).

Helpful Sources :

- Publication Manual of the American Psychological Association, 6th ed. (APA, 2010)

- New Oxford Style Manual (OUP, 2012)

- [Data Citation Awareness Guide](#) (ANDS, 2011)

- [Data Citation: What you Need to Know](#) (ESRC, 2012)

The Librarian could assist researchers, through training and consultation in data citation/referencing methods

Skills needed

- Knowledge to advise and assist researchers on setting up data management plans
- Good knowledge of data sources available in their discipline/subject, e.g. re3data.org
- Keep up-to-date/be familiar with the types of datasets and data formats used by researchers in their subject field/discipline
- Knowledge to advise on data management process
- Knowledge to advise on potential data manipulation/analysis tools used in the discipline/subject, e.g. SPSS, Statistical Tools, 'R', etc.
- Knowledge to advise researchers on metadata schemas in their discipline/subject area
- Excellent skills to design information literacy training pertaining to data
- Ability to advise on current trends, best practice and available options in research publication and dissemination methods, e.g. an understanding of open access publishing

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Skills needed

- Sufficient knowledge to support researchers in complying with the various mandates of funders, including open access requirements, e.g. NRF Statement on Open Access

- Understanding of research impact factors and performance indicators and ability to advise on citation analysis, bibliometrics, etc.

- Understanding of author rights, copyright legislation and IP issues, ethical issues and plagiarism to advise or refer as appropriate

- Ability to advise on the value and use of mobile technologies, Web 2.0 and other communication tools (e.g. Mendeley, Virtual Research Environment) to researchers

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Helpful tools

- **Data Curation Profiles Toolkit** (Purdue) - http://datacurationprofiles.org/ **-** a Tool for conducting data interviews

- **Data Citation Index (Web of Science) -** Counts formal and informal citations of datasets by papers - http://apps.webofknowledge.com/

- **MANTRA** – Free Online Research Data Management Training Course http://datalib.edina.ac.uk/mantra/

- **Essentials 4 Data Support Course** (Netherlands)- free online introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data - http://datasupport.researchdata.nl/en

- **How to Cite Data Sets and Link to Publications** (DCC) http://www.dcc.ac.uk/resources/how-guides/cite-datasets

Helpful Tools (2)

- **How to Track the Impact of Research Data with Metrics** (DCC)
  http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics

- **How to License Research Data** (DCC)
  http://www.dcc.ac.uk/resources/how-guides/license-research-data

- **Impactstory** – a tool to track impact (bibliometrics) http://impactstory.org

- **re3data.org**

  A global registry of research data repositories. A tool for the easy identification of appropriate data repositories to store research data.
  http://www.re3data.org/

- **UP Libguide on RDM -** http://up-za.beta.libguides.com/c.php?g=356288

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Conclusion

data

is the new oil

we need to find it,
extract it,
refine it and
monetize it
David Buckingham

Adapted from:

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Data

Thank you

References

- ***Analyzing and interpreting data.*** Syracuse, NY: Office of Institutional Research and Assessment, Syracuse University, n.d. [Online] available at https://oira.syr.edu/assessment/assesspp/Analyze.htm (Accessed 18 September 2014).

- AUCKLAND, M. 2012. ***Re-skilling for research: an investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers.*** London: RLUK. [Online] available at http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf (Accessed 20 August 2015).

- BALL, A. 2014. ***How to License Research Data.*** (DCC How-to Guides). Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides (Accessed 20 August 2015).

- BALL, A. AND DUKE, M. 2015. ***How to Cite Datasets and Link to Publications.*** (DCC How-to Guides). Edinburgh: Digital Curation Centre. [Online] available at http://www.dcc.ac.uk/resources/how-guides (Accessed 20 August 2015)

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

References

- BALL, A. AND DUKE, M. 2015. *How to Track the Impact of Research Data with Metrics.* (DCC How-to Guides). Edinburgh: Digital Curation Centre. [Online] available at http://www.dcc.ac.uk/resources/how-guides (Accessed 20 August 2015)

- CALLAGHAN, S. et al. 2013. *Connecting data repositories and publishers for data publication.* Presentation delivered on 7 February 2013 at the OpenAIRE Interoperability workshop, University of Minho, held 7-8 February 2013. Braga, Portugal: University of Minho Gualtar Campus. [Online] available at http://openaccess.sdum.uminho.pt/wp-content/uploads/2013/02/7_SarahCallaghan_OpenAIREworkshopUMinho.pdf (Accessed 19 September 2014).

- CHOUDHURY, S. 2014. *Public Institution perspective (Research Library).* Presented at the Digital Media Analysis, Search and Management (DMASM), 2014. [Online] available at http://dataconservancy.org/wp-content/uploads/2014/03/DC_DMASM_2014.pdf (Accessed 24 September 2014).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

References

- CORTI, L. et al. 2014. ***Managing and sharing research data: a guide to good practice.*** Los Angeles: SAGE.

- ***Data Citation Awareness guide.*** Australian National Data Service [Online] available at http://www.ands.org.au/guides/data-citation-awareness.html (Accessed 20 August 2015)

- ***Data Management Planning Tool (DMPTool).*** Oakland, CA: University of California, California Digital Library, 2015. [Online] available at https://dmptool.org/ (Accessed 20 August 2015)

- ***DMPOnline tool***, 2015. Edinburgh, UK: Digital Curation Centre. [Online] available at https://dmponline.dcc.ac.uk (Accessed 20 August 2015)

- ***Essentials 4 Data Support Course.*** [sl.:sn.] [Online] available at http://datasupport.researchdata.nl/en (Accessed 20 August 2015).

- ***Impactstory.*** [n.d.] [sl.: s.n.]. [Online] available at https://impactstory.org/ (Accessed 20 August 2015).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

References

- ***MANTRA: Free Online Research Data Management Training Course.*** Edinburgh, UK: University of Edinburgh. [Online] available at http://datalib.edina.ac.uk/mantra/ (Accessed 20 August 2015)
- MARQUES, D. 2013. Research data driving new services. ***Elsevier Library Connect***, 25 February 2013. [Online] available at http://libraryconnect.elsevier.com/articles/best-practices/2013-02/research-data-driving-new-services (Accessed 5 October 2014)
- MARTIN, E. AND BALLARD, G. 2010. ***Data management best practices and standards for Biodiversity data applicable to Bird Monitoring Data.*** U.S. North American Bird Conservation Initiative Monitoring Subcommittee. [Online] available at http://www.nabci-us.org/ (Accessed 24 September 2014).
- ***New Oxford Style Manual.*** 2nd ed. Oxford: Oxford University Press, 2012
- ***The Oxford Dictionary.*** [sl.]: Oxford University Press, 2014. [Online] available at http://www.oxforddictionaries.com/us/ (Accessed 16 September 2014).

References

- ***Publication Manual of the American Psychological Association***, 6th ed. Washington, DC: American Psychological Association, 2010.

- ***re3data.org: registry of research data repositories***, n.d. [sl.: s.n.]. [Online] available at http://www.re3data.org/ (Accessed 20 August 2015).

- ***Research Data Services, University of Wisconsin-Madison***. Madison, WI: University of Wisconsin Madison, 201. [Online] available at http://researchdata.wisc.edu/manage-your-data/data-backup-and-integrity/ (Accessed 24 September 2014).

- ***Responsible conduct of research.*** DeKalb, Illinois: Northern Illinois University Faculty Development and Instructional Design Center, n.d. [Online] available at: http://ori.dhhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html. (Accessed: 16 September 2014).

- ***Standard naming conventions for electronic records: the rules.*** Edinburgh: University of Edinburgh, 2015. [Online] available at http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/RMprojects/PP/FileNameRules/Rules.htm (Accessed 20 August 2015).

References

- ***Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research.*** Pretoria: National Research Foundation, 2015. [Online] available at http://www.nrf.ac.za/media-room/news/statement-open-access-research-publications-national-research-foundation-nrf-funded (Accessed 20 August 2015).

- United States Environmental Protection Agency (US EPA). 2002*.* ***Guidance on Environmental Data Verification and Data Validation: EPA QA/G-8.*** Washington, DC: Environmental Protection Agency. [Online] available at http://www.epa.gov/QUALITY/qs-docs/g8-final.pdf (Accessed 24 September 2014).

- ***USGS Data Management.*** [Online] available at http://www.usgs.gov/datamanagement/describe/metadata.php (Accessed 19 August 2014).

- YUAN, X. 2014. ***Visualization and visual analytics.*** Presentation 0n 12 June 2014 at CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, Beijing, China, 4-20 June 2014.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA