

Survey of Research Data Management Practices at the University of Pretoria

Undertaken by the Department of Library Services in
order to improve research practices at the University

Unisa Library Open Access Event
25 October 2011

Dr Heila Pienaar, UP Library Services,
University of Pretoria, South Africa
http://www.ais.up.ac.za/profile/heila_pienaar/index.htm

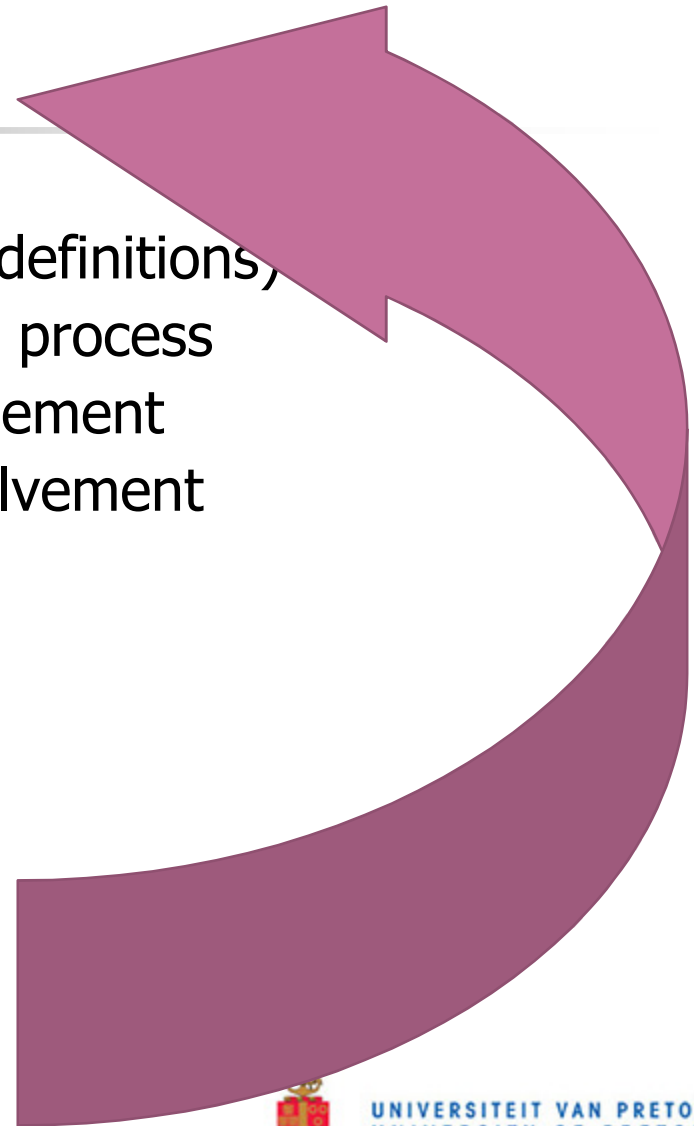


UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



Content

- Context: UP Library e-Strategy
- Data curation / management (definitions)
- Data management – concepts, process
- Levels of research data management
- Rationale for the Library's involvement
- Survey research methodology
- Findings
- Top requirements for services
- National initiatives
- Recommendations
- Further actions
- Anecdotes from the survey!

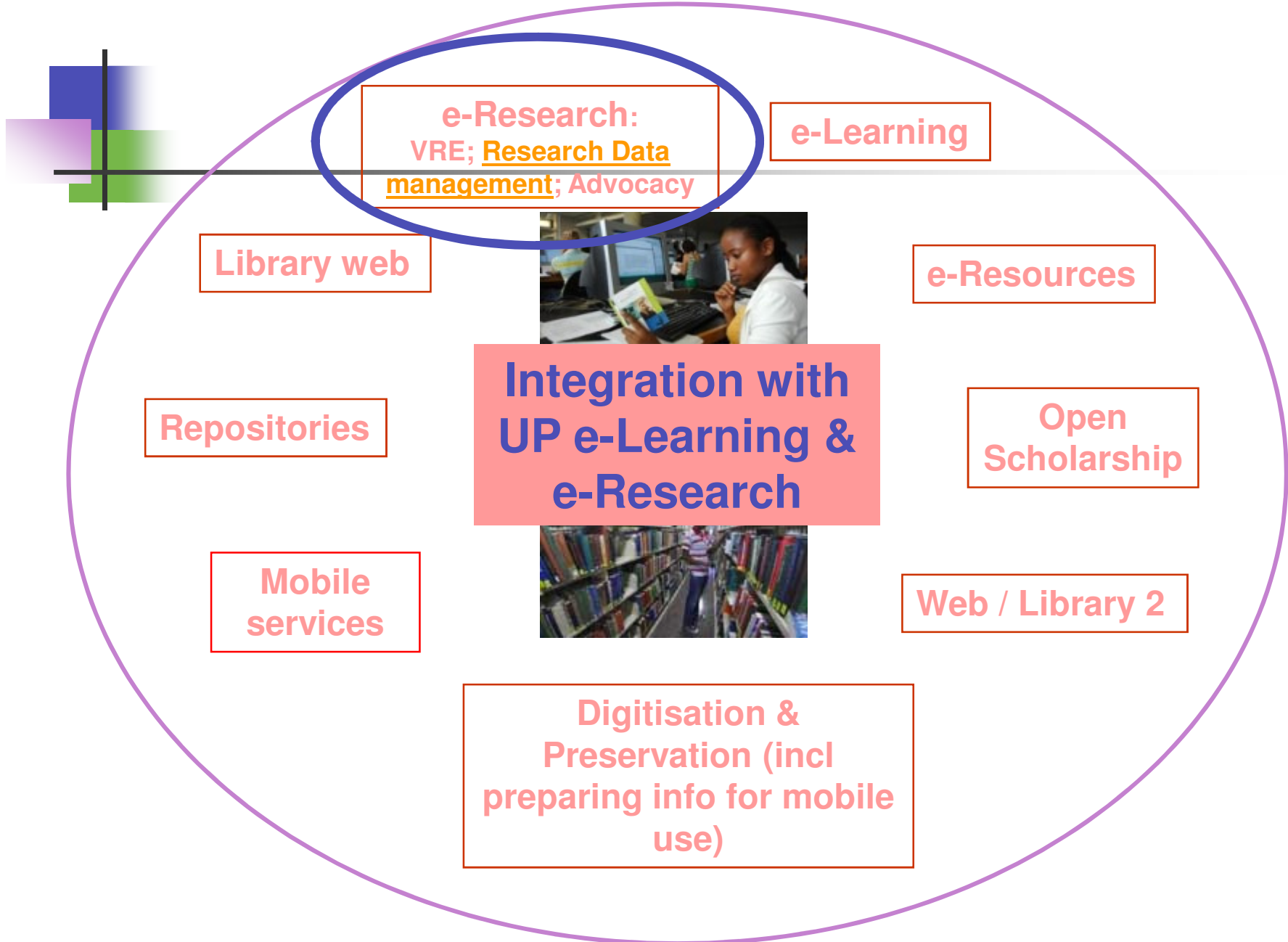




Anecdotes from the survey

- Department of Chemistry: Data is very vulnerable because of the tendency of chemicals to burn
- Faculty of Engineering: When students graduate they download their data on DVD's etc and give it in a paper file to the supervisor. No backups.
- Faculty of Natural and Agricultural Sciences: A student was hijacked and all data stolen, plus lab book. Data could not be recovered.

e-Information Strategy & projects 2011



Data & data curation / management (definitions)



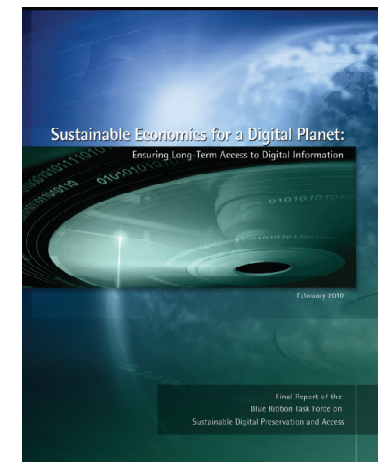
- A relatively new discipline with many different definitions
- **Research data:** Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results <http://www.ed.ac.uk/is/data-management>
- **Data curation:** the curation of records or measurements of information ("data"). Those scientific measurements or records ("data") are further distinguished from the computer science meaning of "data" to refer to any type of digitally encoded information
http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean
- **Digital curation:** the selection, preservation, maintenance, collection and archiving of digital assets
http://en.wikipedia.org/wiki/Digital_curation

Blue Ribbon Task Force on Sustainable Digital Preservation and Access

The Task Force's view on research data:

"There is a remarkable growth of data-intensive research in all knowledge domains. In most fields, there is high recognition of the benefits of preserving research data for various purposes and lengths of time. But there are few robust systems for making decisions about what to preserve; and there is often a lack of coordination of roles, responsibilities, and funding sources among those best positioned to preserve data (researchers) and the preservation infrastructure (curation and archiving services) that should support them. Research and education institutions, professional societies, archives, researchers, and the funding agencies that support data creation all have leading roles to play in creating sustainable preservation strategies"

(http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf)





Terminology (Blue Ribbon Task Force)

- The **terminology** for digital materials and preservation processes **varies among stakeholder communities**.
- As a rule members of the **scientific community** refer to digital materials as **data**; further, they refer to activities that enable use and long-term accessibility as **curation and archiving**, which taken together, are called **stewardship**.
- In **cultural** domains and the **humanities**, digital materials are more often referred to as **content**, and the activities that ensure their long-term availability are called **preservation and access**.



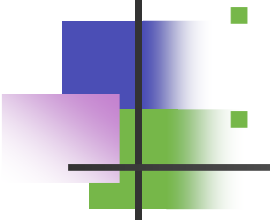
Research data management

Is not only:

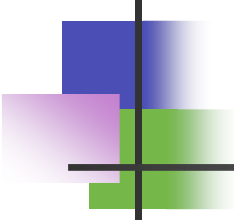
Data archiving OR

Data backups

Data Management concepts, process ...

- 
- **Data Ownership** This pertains to who has the legal rights to the data and who retains the data after the project is completed.
 - **Data Collection** This pertains to collecting project data in a consistent, systematic manner (i.e., reliability) and establishing an ongoing system for evaluating and recording changes to the project protocol (i.e., validity).
 - **Data Storage** This concerns the amount of data that should be stored -- enough so that project results can be reconstructed.
 - **Data Protection** This relates to protecting written and electronic data from physical damage and protecting data integrity, including damage from tampering or theft.
 - **Data Retention** This refers to the length of time one needs to keep the project data according to the sponsor's or funder's guidelines. It also includes secure destruction of data.
 - **Data Analysis** This pertains to how raw data are chosen, evaluated, and interpreted into meaningful and significant conclusions that other researchers and the public can understand and use.
 - **Data Sharing** This concerns how project data and research results are disseminated to other researchers and the general public, and when data should not be shared.
 - **Data Reporting** This pertains to the publication of conclusive findings, both positive and negative, after the project is completed. (Steneck, 2004) <http://ori.dhhs.gov/education/products/clinicaltools/data.pdf>

Why manage research data?

- 
- **Data management is one of the essential areas of responsible conduct of research.**
 - Before starting a new research project, the researchers and or the research teams must address issues related to data management.
 - By managing your data you will:
 - Meet funding body grant requirements.
 - Ensure research integrity and replication.
 - Ensure research data and records are accurate, complete, authentic and reliable.
 - Increase your research efficiency.
 - Save time and resources in the long run.
 - Enhance data security and minimise the risk of data loss.
 - Prevent duplication of effort by enabling others to use your data.
 - Comply with practices conducted in industry and commerce.

<http://www.ed.ac.uk/is/data-management>

Levels of research data management




- International e.g. World Data Centre on Climate
- National e.g. Very Large Database Initiative (DST: CHPC, Meraka, CSIR); NeDICC (Network of Distributed Data & Information Curation Centres) Initiative
- Campus e.g. repositories for open access



Rationale for the Library's involvement

- “Thus with the experience gained from **traditional** cataloguing, indexing and organizational **skills** coupled to those acquired in developing, establishing and maintaining institutional **repositories**, the time is ripe for academic librarians to explore their role as **data curators**”
- Data Curation and Libraries: Short-Term Developments, Long-Term Prospects
http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean
- A new role for academic librarians: data curation
<http://www.era.lib.ed.ac.uk/handle/1842/3207>

Resources needed

- 
- People / expertise / systems involved in data management and sharing may include:

project director designing research (incl. research data management)

research staff collecting, processing and analysing data

external / internal contractors involved in data collection, data entry, processing or analysis

support staff managing and administering research and research funding

institutional IT services staff providing & advising on formats, repositories, data storage, data archiving and back-up services

external data centres or web services archives who facilitate data sharing

meta-data editors: data description, annotation and contextual information



Survey research methodology

- Fifty-two interviews were conducted by 15 information specialists from the relevant Faculty Libraries over the period October 2009 – March 2010
- Each Faculty's Research Committee was requested by the Vice Principal: Research and Postgraduate Studies to identify up to three researchers to take part in the survey
- Each researcher also identified at least one post-graduate student who could participate in the survey.
- The information specialists received formal training in interview techniques.
- Interviews were conducted according to a semi-structured interview framework.
- **Limitation:** Results cannot be generalised to researchers not included in this study



Distribution of interview respondents

	<u>Faculty</u>	<u>Academic staff</u>	<u>Post-graduate students</u> (some are also academic staff)	Total
1	Theology	2	2	4
2	Humanities	4	4	8
3	Education	3	1	4
4	Law	5	-	5
5	Economic & Management Sciences	8	?	8
6	Health Sciences	3	3	6
7	Veterinary Sciences	3	3	6
8	Natural and Agricultural Sciences	5	2	7
9	Engineering, Built Environment & Information Technology	2	2	4
	Total	35	17	52



Findings

- The general trends of findings are given, using six major categories:
 - funding,
 - data collection,
 - processing of data,
 - publishing,
 - support



Funding

- Funding. This part of the interview was included in order to better understand how researchers think about data at the early stage of applying for funding and how well they are aware of their funders' requirements in terms of data sharing and archiving.
- General trend: It depends on the funding agency proposal requirements and in most cases there is no need for data management or data sharing plans.



Data Collection

- Data collection. This aspect was discussed in order to learn about the different ways in which data are collected and captured, the **different types of formats and sizes** as well as the usefulness of these data to others.
- General trend: UP researchers make use of a wide variety of data collection methods and use both primary and secondary data. Both 'soft' and 'hard' data collection methods are used by all the Faculties. Data sets are often small



Processing of data

- Processing of data. In this portion of the interview the aim was to understand how researchers **store data securely**.
- General trend: Ad hoc storage of data, both on paper and electronically, is the norm. A few servers are available for data storage but in general the onus rests on the individual or department on how and where data is stored



Publishing

- Data publication. This part of the interview was included to see how researchers **publish their data**, if they do, and to explore the reasons behind not publishing data at all.
- General trend: In general raw data is not published for other researchers to use, and it is also not seen as necessary to do so.



Support

- Support. This section of the interview was designed to learn about the support researchers receive **to manage their data** and where they turn to for help when they encounter problems.
- General trend: Support for research as an activity is good throughout the university (faculty, departments, research support). But there is a lack of support with regard to the storage of data (physical and electronic).



Top requirements for services

- Top requirements for services. At the end of the interview challenges and concerns in terms of managing their data, were discussed with interviewees and they were asked to suggest services that could help them do their work more effectively.
- General trend:
 - The top requirement is a central UP server or repository that is **easy to use** with **good security**.
 - There is also a need for **physical storage space**.
 - The biggest worry of academic staff is lack of sufficient time and
 - lack of support for research by the UP Executive.



National initiatives

- Very Large Data Base (VLDB) – mandated to the Centre for High Performance Computing (CHPC) by DST
- UP Library and the VLDB organised a Library Directors' workshop to help identify research data management needs of SA universities



Recommendations

- It can be safely said that 'research data management' does **not exist** in any formal manner (with the exception of one or two departments) at the University of Pretoria
- The **Very Large Database initiative** from the Department of Science and Technology should be investigated to see if it would support UP's research data management needs
- A formal staff position of '**research data manager**' is needed whether UP makes use of an external or internal system / repository or not. Such a position is necessary to drive the research data management endeavor



Acknowledgement

This survey and report structure is based to a large extent on the “Findings of the scoping study interviews and the research data management workshop. Scoping digital repository services for research data management. A Project of the Office of the Director of IT www.ict.ox.ac.uk/odit/projects/digitalrepository/” by Luis Martinez-Urbe (luis.martinez-uribe@oerc.ox.ac.uk), Digital Repositories Research Co-ordinator, University of Oxford, UK



Further actions:

- Library has allocated a staff position for UP research data manager
- Presented the report at the UP Senate Research Committee
- Requested by prof Robin Crewe, previous Vice Principal: Research and Postgraduate Studies, to identify a Technical solution
- Decided at first meeting between Library, IT & Research support to evaluate the maturity of UP to manage research data

Example of maturity model (draft)

Maturity of the University of Pretoria to implement research data management

Maturity ↗ Checklist for research data** ↓	0 = Non-existent	1 = Initial / Ad hoc	2 = Repeatable but intuitive	3 = Defined process	4 = Managed & measurable	5 = Optimised
Copyright protection of data is clear						
Ownership of copyright & IP is clear						
Ethical requirements are applied						
Durable & acceptable data formats are used						
Digital data storage & backups are in place						

*Based on the CobiT framework generic maturity model: http://www.ee.kth.se/php/modules/publications/reports/2007/IR-EE-ICS_2007_026.pdf

**Monash University Library Research Data Planning Checklist: <http://www.researchdata.monash.edu/resources/datahdrchecklist.doc>



Further actions (continued)

- 'Proposal on the implementation of research data management at UP' accepted September 2011 by UP Research Computing Committee
- Role description and advert for 'UP Research Data Manager' to be finalised (appointment in 2012)
- Pilot study 1: Research data to be included as part of a Department's electronic theses & dissertations
- Pilot study 2: Urgent request by Faculty of Health Sciences to support their research data management process



The end, for now

- Comments, Questions???