Survey of research data management practices at the University of Pretoria, South Africa: October 2009 – March 2010

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

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

22nd International CODATA Conference
24-27 October 2010
Cape Town, South Africa
http://www.codata2010.com/
Content

- Data curation / management (definitions)
- Data management – concepts, process
- Levels of research data management
- Rationale for the Library’s involvement
- Research methodology
- Findings
- Top requirements for services
- National initiatives
- Recommendations
- Further actions
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](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](http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean)
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](http://ori.dhhs.gov/education/products/clinicaltools/data.pdf)
UK Digital Curation Centre’s Lifecycle Model

http://www.dcc.ac.uk/resources/curation-lifecycle-model
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 / 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
  [Link](http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1027&context=lib_dean)

- A new role for academic librarians: data curation
  [Link](http://www.era.lib.ed.ac.uk/handle/1842/3207)
Research methodology

- **Fifty-two interviews** were conducted by 15 information specialists from the relevant Faculty Libraries.
- 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

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Academic staff</th>
<th>Post-graduate students (some are also academic staff)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theology</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Humanities</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Education</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Law</td>
<td>5</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Economic &amp; Management Sciences</td>
<td>8</td>
<td>?</td>
<td>8</td>
</tr>
<tr>
<td>Health Sciences</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Veterinary Sciences</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Natural and Agricultural Sciences</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Engineering, Built Environment &amp; Information Technology</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>35</strong></td>
<td><strong>17</strong></td>
<td><strong>52</strong></td>
</tr>
</tbody>
</table>
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 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-Uribe (luis.martinez-uribe@oerc.ox.ac.uk), Digital Repositories Research Co-ordinator, University of Oxford, UK


Further actions:

- Library is requesting a new staff position for research data manager.
- Presented the report at the UP Senate Research Committee.
- Requested by prof Robin Crewe, 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**

<table>
<thead>
<tr>
<th>Maturity Checklist for research data</th>
<th>0 = Non-existent</th>
<th>1 = Initial / Ad hoc</th>
<th>2 = Repeatable but intuitive</th>
<th>3 = Defined process</th>
<th>4 = Managed &amp; measurable</th>
<th>5 = Optimised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright protection of data is clear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ownership of copyright &amp; IP is clear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethical requirements are applied</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durable &amp; acceptable data formats are used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital data storage &amp; backups are in place</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


The end, for now

- Comments, Questions???