Data Management Plans (DMPs)
Writing DMPs using DMPonline
07 September 2016

Erika Mias
Digital Curation Officer
UCT Libraries, DLS
University of Cape Town
Overview
What is a DMP?

• “…a formal document that describes the data produced in the course of a research project. [It also] outlines the data management strategies that will be implemented both during the active phase of the research project and after the project ends.”

- Sarah Jones (DCC)
Why create DMPs?

- Mandatory funder requirement
- Mandatory Institutional requirement
- Part of good research practice: your research data still needs management throughout the research lifecycle.
- Well-managed data allows for:
  
  verification or refinement of published research results,
  
  reduces the potential for scientific fraud,
  
  promotes new research through the use of existing data,
  
  provides resources for training new researchers and discourages unintentional redundancy in research. By planning for data management, these benefits are more likely to be realized.
Basic DMP questions:

1. What data will you collect or create?
2. How will the data be collected or created?
3. What documentation and metadata will accompany the data?
4. How will you manage any ethical issues?
5. How will you manage copyright and Intellectual Property Rights (IPR) issues?
6. How will the data be stored and backed up during the research?
7. How will you manage access and security?
8. Which data should be retained, shared, and/or preserved?
9. What is the long-term preservation plan for the dataset?
10. How will you share the data?
11. Are any restrictions on data sharing required?
Research Data Management (RDM) at Digital Library Services (DLS)
Data Management Planning

RDM at Digital Library Services

- DLS encourage researchers to **manage their research data**:  
  - **rewards of sharing** research data
- DLS assist researchers with data management planning:  
  - **DMPonline**, **policies** and **guidelines**
- DLS provide tools and services to support researchers in depositing, preserving and sharing their data:  
  - Implementation of repository infrastructures
- [www.digitalservices.lib.uct.ac.za](http://www.digitalservices.lib.uct.ac.za)
Welcome.

DMPonline has been developed by the Digital Curation Centre to help you write data management plans.

Screencast on how to use DMPonline

Sign in

Email address *

Password *

Forgot your password?

Remember me

Sign in

Sign up

New to DMPonline? Sign up today.
Why have a instance UCT DMPonline

• Data is stored and managed locally
• Loading of customised templates with unique institutional-specific guidance
• Administrative access for departmental data managers

DMPonline demonstration follow at:

dmp.lib.uct.ac.za

Watch the video for more info on the admin interface functionality, to set up the account contact Erika Mias.

07 September 2016
## UCT DMPonline roadmap beta release

### Roadmap Themes

<table>
<thead>
<tr>
<th>Theme</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability improvements</td>
<td>Features that make the tool more usable for researchers and research administrators.</td>
</tr>
<tr>
<td>UCT Community DMP Support</td>
<td>Institution templates, outreach and training, collaboration with broader research communities.</td>
</tr>
<tr>
<td>Systems integration</td>
<td>Converis integration, OSF integration, metadata harvesting etc..</td>
</tr>
<tr>
<td>Planned Updates</td>
<td>Scheduled template uploads, e.g. funder templates, software updates etc..</td>
</tr>
<tr>
<td>Functional Requirement Update</td>
<td>Software upgrades, customizations requiring “hard-coding” etc...</td>
</tr>
<tr>
<td>Active Maintenance</td>
<td>Bug fixes, customer support</td>
</tr>
</tbody>
</table>
Data Management Planning

UCT DMPonline roadmap beta release

Pre-Release (October 2015)
- Bug fix: shared with user community
- DMPonline V4.04 installed in DEV mode

Pre-Release (November 2015)
- NERC template
- MRC-UK template
- NSF- Bio template
- Guidance Enhancement: UCT links*

Pre-Release (January 2016)
- CRC template (v.01)
- UCT template (v.01)

Release 1 (February 2016)
- DMPonline migrated to ICTS
- DMPonline Live, DLS Launch 4 Feb
- UCT template V.01**
- Bug fix: NEDIC DMPonline installation

Release 1 (April 2016)
- UCT template (v.02)
- EU Horizon 2020 template

Release 1 (April 2016)
- UCT template V 02.1 ***
- DCC default replaced with UCT template****
- DMPonline @UCT webpage updates
- Wellcome Trust Template

07 September 2016
Data Management Planning

**UCT DMPonline roadmap beta release**

- **Release 2 (June 2016)**
  - DMPonline @ UCT case study presentation
  - NEDICC

- **Release 2 (July 2016)**
  - DMPonline NEDICC collaboration
  - ****

- **Release 2 (August 2016)**
  - DMPonline NEDICC roadmap release

- **Release 2 (September 2016)**
  - SSO service working

- **Release 2 (October - December 2016)**
  - Gates Foundation Template

- **Release 2 (September 2016)**
  - DMPonline NEDICC template

- **Release 2 (September 2016)**
  - DMPonline NEDICC metadata harvest test

- **Release 2 (September 2016)**
  - NIH Template

- **Release 2 (September 2016)**
  - DMPonline training data mangers

07 September 2016
Best practices for working with your data
Think like a digital curator

File management

- File management practices help you **identify, locate and use** your data effectively
- Good file management helps others to understand, collaborate and/or reuse your data effectively

- Well managed files are:
  - distinguishable
  - easy to locate and browse
  - not easily overwritten or deleted
  - easy to collaborate with
  - easy to work with (open formats)
Think like a digital curator

File management

• File management practices help you **identify, locate and use** your data effectively
• Good file management helps others to understand, collaborate and/or reuse your data effectively

• Well managed files are:
  distinguishable
easy to locate and browse
not easily overwritten or deleted
easy to collaborate with
easy to work with (open formats)
File naming

- **Three criteria to assist with naming files:**
  - Organisation
  - Context
  - **Consistency**

- Elements to consider when naming files:
  - version numbers
  - creation / publication date
  - creator’s name / group name
  - content description
  - project number

Image Credit: Cliparts
File naming rules

- Always consider **scalability** when naming files
e.g., 001 vs 01
- Keep file names **short & relevant**
- **Don’t** use special characters or spaces
  underscores instead of full-stops
  dashes instead of spaces
- See the University of Edinburgh resource on [file naming conventions](#).

Renaming files

- Operating systems have built-in tools for **batch renaming**
- Software tools available for batch renaming files
  [name changer](#)

07 September 2016
File versioning

Always record changes to your data files, even if it seems unnecessary!

• Indicating major version changes
  V1, V2, etc
• Indicating minor version changes
  V1.1… V1.2… etc
• Avoid using labels
  ‘final’, ‘final3, ‘draft’
• Version control software
  Subversion, TortoiseSVN
File formats

- File formats **encode information about a file** that enable it to be **recognised** by a computer program or application.
- File formats are indicated in the filename by an **extension** that follows a full-stop.
  - jpg, docx, pdf
- **Proprietary** vs **Open** file formats
  - Proprietary file formats can only be opened by the software used to create the file.
  - Open file formats are openly available and can be recognised by a number of applications.
- **File format obsolescence**
  - Changes in technology
  - Updates of software
  - Collaboration across platforms

*Convert to open formats and save along with proprietary files...*
File formats continued...

- **Migration vs Normalisation**
  - both involve converting files from one format to another (typically preservation-friendly, open formats)
  - Migration refers to the conversion of files when the file format is at risk of obsolescence
  - Normalisation is the practice of converting file formats upon acquisition for long-term preservation

Always a good idea to **normalise** your files to ensure preservation and avoid migration!

- DPC - File formats and standards
- Stanford University - Best Practice for file formats
- Archivematica - Format Policy Registry
- DCC - Open source software and open standards
- Open Data Handbook - File formats
Data transformation

- Involves changing the actual data (not file format)
  - de-identification, anonymisation
  - Converting qualitative data into quantitative data
  - Converting numerical data into bar graphs

- Data transformation enables further analysis of the data collected
What is metadata?

Image Credit: Cliparts
Think like a digital curator

What is metadata

“A set of data that describes and gives information about other data”. (source: Oxford Dictionaries)

“Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information. “ NISO, 2004

- DCC list of disciplinary metadata
- UCT metadata entry guidelines

Document your data while you’re creating it so that it is easy to understand and use later on...
Documentation

● **Readme Files**
  ○ Advantages of creating well structured Readme files: (See Daniel Beck’s excellent [Readme checklist and presentation](#))
  ○ **Helps the reader to identify, evaluate, use and engage with your project.**
  ○ Increasing requirement of data repositories and funders to submit a readme file when depositing in a data repository.

● **Codebooks and Laboratory Notebooks**
  ○ Raw data such as these also need to be managed effectively and preserved where possible- NB for reproducibility. (Researcher interview: [Shaun Bevan](#))
Storing and backing up your data

- **Storage**
  - From the outset, think about how much storage you require
  - Think about who needs to access your data and how that affects your storage location
  - Include costs of storage in your DMP and funding applications
  - Network drives highly recommended for storing and accessing master copies
    - UCT eResearch data storage services

- **Backup**
  - Find out about your network provider’s backup services
  - Set up your own backup workflow…
    - daily / weekly / monthly
    - 2 - 3 copies, different locations
    - Incremental vs. Full
    - Cloud vs. Local

- **Security**
  - Who needs access? How will you control access?
  - Sensitive data and encryption

researcher interview on file management and security: Natalia Calanzani
Extra work!?

- **Data transformation can lead to unforeseen uses** in other disciplines. New use cases = more citations and recognition for your work.

- **Data management** makes it easier to preserve and archive data. Makes it easier to plan your research project to meet requirements. Increases the likelihood of reproducing your results and validating your research. Eases the transition for new project members and collaborators.

Source: [Cliparts](#)
Other Resources

- **Research Data Management Tutorials**
  - [Mantra](#)
  - [Leeds University](#)
  - [Research Data Management and Sharing](#) (Coursera)

- **DMPonline tutorials**
  - [EUDAT presentation on writing a DMP](#)

- **Metadata schemas and disciplinary metadata**
  - [Overview of metadata types](#)
  - [Disciplinary metadata](#)
  - [Metadata standards](#)

- **Open Research**
  - [SPARC](#)
  - [Why Open Research](#)

- **Researcher Interviews**
  - [Odum Institute](#) interviews researchers on why RDM is important.

*(On a less serious note: “A Data Sharing and Management Snafu in 3 Parts”)*

07 September 2016
Thank You
Questions?

Follow us on Twitter: Digital@UCT and DMPonlineUCT