

## **DataTrain Archaeology Module 2**

### **Data Lifecycles and Management Plans**

#### **Presentation Notes**

**Written by Lindsay Lloyd-Smith**

**(July 2011)**

#### **Slide 1      Data Lifecycles and Management Plans**

There are some general concepts in data management which might sound really boring but actually are quite useful to appreciate.

As somewhere to start from these concepts help in organising your own, individual, ways you look after your own research data, particularly when it comes to making your data understandable to others in the future.

---

#### **Slide 2      A Bit of Theory**

That said, though, you'll be glad to hear that the 'theory' bit will be kept to a minimum.

One important aspect to cover is the jargon used by people who run digital repositories. Most researchers probably won't have cause to use this jargon in their work, but it is useful to know what they're talking about when we come to deposit digital data in a repository.

Another benefit of thinking of research data in terms of a Lifecycle Model is that this concept forms the basis of Data Management Plans which are now a requirement for large research grant applications. The Arts and Humanities Research Council uses the term "Technical Appendix", but it amounts to same thing.

And at the end of this session you will begin to draft your own Data Management Plan for your post-graduate research project.



### **Slide 3      Data Lifecycle**

A data lifecycle model divides the research process into a number of tasks: project planning; data collection; data analysis; data distribution and archiving; data discovery and re-use; leading to data re-analysis, and so on and so forth.

Indeed many post-graduate research projects are based on re-analysis of existing data, which in turn produces new digital data which has to be looked after and archived. The last picture here is of a PhD student, who hadn't seen the light of day for many weeks, printing out his thesis.

Thinking of data in terms of a life cycle helps to:

- Identify and define the different tasks involved in making a realistic and effective plan to look after the data during the course of the project;
  - Plan ahead and define what the out come of the project will be;
  - Think about where the data - both physical and digital - will end up;
  - Think about what formats the data will need to be in;
  - And what documentation we will need to provide so that the data are re-usable in the future.
-

#### **Slide 4      Digital Data Curation Centre Lifecycle Model**

A data lifecycle model can be applied to all types of work in all disciplines, both research and professional.

This is the generic Data Lifecycle Model produced by the Digital Curation Centre, a government funded centre that advises on all things digital. I know it looks quite mad and overly complicated.

It is designed for large projects with many people working at different levels with different responsibilities. Specifically this model is to help project managers allocate tasks and responsibilities to project members to make sure everybody knows who is doing what, when, and why.

The different circles denote the timing of different tasks, highlighting those which take place at particular times during the course of the project and those which are continual.

The circular shape of the model also highlights the important fact that the whole point of preserving digital data is for it to be accessed and re-used in the future.

Archaeologists appreciate this point very well: we destroy data to create data, and much new research is based wholly or in part on re-analysis of old data sets.

The main reason for showing this crazy model, though, is that it introduces a number of terms that are common in data management that you will come across later in your careers when talking with digital people, for example at the Archaeology Data Service. So, it is useful to cover quickly some definitions before moving on.

---

## Slide 5      Some Jargon and Definitions

Jargon has been kept to a minimum in these slides and exercises, but it is important to be familiar with a few terms:

Digital Data Capture	Data that are derived from a physical data object e.g. inputted measurement of artefacts, scans of drawings, etc.
Digital Data Creation	Data that are 'Born Digital' and not derived from a physical data object, e.g. digital photographs, geophysical survey data, scientific analysis readouts, etc.
Documentation	Explains how data were created or digitised, what the data mean, what their content and structure are, and any manipulations that may have taken place.
Metadata	<p>"<i>Data about Data</i>". Standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection.</p> <p>From an archivist's and librarian's point of view <b>metadata</b> are also absolutely crucial when it comes to <b>retrieval of digital materials</b>, this is because it the only way of finding material through searches of the 1000s of files held in the repository.</p>
Data storage / Back-up	System used for looking after digital data during the life of a project. <b>Back-up is NOT preservation.</b>
Ingest	Process by which digital data are archived by a digital depository.
Digital Preservation	Long term archiving of digital data so that it will be accessible in the future.
Long Term	Period over which changing technologies, formats, media impact upon the access to and use of digital resources.
Migration	Transfer of digital resources from one hardware/media and software/file format generation to the next.
Emulation	Techniques for imitating obsolete computer systems to retrieve digital data. (You might hear computer people talk about this but it is unlikely you will need to worry about it).

---

## **Slide 6      Data Lifecycles and Data Management Plans - overview**

A simplified data life cycle might look something like this.

And all a Data Management Plan is is a project document that works through each of these stages around the lifecycle and answers each of these questions:

- What types of digital data will I produce?
- How will I organise the data? (file structure/naming, formats, software)
- At what points during the project, and how, will the data be evaluated?
- What data will be kept and what deleted?

At this point some data may be re-cycled, re-analysed, and feed back into the management process. Or, new data may be required.

The remaining parts of the management plan address the questions:

- What data will be deposited and where?
- Who will be interested in re-using the data?

A possible analogy to a Data Management Plan is a Risk Assessment Form – which is completed before we carry out any fieldwork and where we think through all the possible things that could go wrong, the procedures in place to mitigate the risks, and finally the actions that will be taken if something does go wrong.

In a similar way the purpose of a Data Management Plan is to think through and explicitly define in black and white the data that will be created, how it will be looked after, and what data will end up where.

---

## **Slide 7      Data Lifecycles and Data Management Plans - 1**

Not only do we need to think of all the general types of digital data, for example digital images, audio recordings, but we need also to think about the specific file formats that will be used.

For example for digital images we need to think about the properties of the files: colour range used, resolution, and size of the images.

For all types of data we need to try to estimate how many individual files will be produced. For example if we are studying five cemeteries each with 200 burials, what types and how many digital files will be produced for each burial?

We should also try to estimate the size of the digital data. This very hard to do at the beginning of a project, but try to think about the size of a typical individual file and scale up.

All these points will be revised during the course of the project. In the first instance, the important thing is to have defined your research data, and file types, and have begun to envisage how large your project archive will end up.

---

## **Slide 8      Data Lifecycles and Data Management Plans - 2**

The second part of a data management plan is where you describe in writing exactly how you will organise and look after your data.

Here you describe the structure and naming of project folders and files, the formats and software that you will use.

Here also you describe how and where the different types of data will be documented, for example whether you are following standardised terminology to describe the data, and where contextual information of data will be provided (in the file name or in a separate file, e.g. spreadsheet or text file).

---

## **Slide 9      Data Lifecycles and Data Management Plans - 3**

It is important to state when and by what criteria in the life of research project the data management plan will be evaluated.

Evaluation often takes place after the initial data sets have been collected and processed. For post-graduate research projects this usually means at the end of the first year or after the first field season of data collection.

If a Second Year Report is written as part of the doctoral program then a revised and updated Data Management Plan can be included then.

---

## Slide 10      Data Lifecycles and Data Management Plans - 4

Part of the data selection process is defining what data will be included in the final project publication – in your case your doctoral theses – and what data will only be found in the supplementary appendices.

- Are you required / intending to submit an E-thesis that will be available online?
  - Will the appendices be included with the thesis?
  - Will a core data set be deposited with a digital repository?
- 

## Slide 11      Data Lifecycles and Data Management Plans – 5

When planning to deposit research data it is useful to ask yourself:

- Who might potentially be interested in re-using the data you have produced?

Thinking about who the potential interested parties are often helps in coming up with the best plan for disseminating the results of your research and the supporting digital data.

Interested parties often include: research communities (sub-discipline, chronological, geographical), museums, local community groups where the research was carried out, and the general public and media.

---

## Slide 12      Data Lifecycles and Data Management Plans – 6

This last issue brings us back to the beginning.

If we intend to deposit the digital data that we produce during the course of our research, we have to make sure – **before we even start collecting it** – that we will have the authority to archive it somewhere.

This means seeking and obtaining permissions from the official bodies (e.g. museums or professional excavation units), or individual researchers or local community groups who may have provided access to, or be an integral part of, the research data.

---

### **Slide 13      Exercise 2: Drafting a Data Management Plan...in reverse order**

One way to start drafting a plan for how you will look after your own research data is to think of it in reverse order:

- What do you want or are required to do with your research data after completing your PhD?

By working towards this goal from the outset you will organise your data with this in mind as you progress with your research.

Building upon what you wrote on the first exercise form, the questions in this second short exercise lead you in reverse order into drafting your data management plan.

Spend 5-10 minutes writing a sentence or two in answer to each of these questions:

- What do you want / have to do with your research data after completing your PhD?
  - Are any of the data sensitive?
  - Who 'owns' your research data and will you have authority to archive it?
  - Describe any copyrighted material you plan to include in your thesis, e.g. images.
  - Will you be working with any unusual or large file formats?
  - Describe the core data set of your research.
-

## Slide 14 Acknowledgements

Module 2 Data Lifecycles and Management Plans

Written by Lindsay Lloyd-Smith (2011)

### *Acknowledgements*

This material was created by the JISC-funded DataTrain Project based at the Cambridge University Library.

Project Manager: Elin Stageland (Cambridge University Library)

Project advisors: Stuart Jeffrey (Archaeology Data Service), Sian Lazar (Department of Anthropology, Cambridge University), Irene Peano (DataTrain Project Officer: Social Anthropology), Cameron Petrie (Department of Archaeology, Cambridge University), Grant Young (Cambridge University Library), and Anna Collins (DSpace@Cambridge Research Data and Digital Curation Officer).

### *Image credits:*

Slide 3 images from left to right:

*The Haddon Library* courtesy of Cameron Petrie, Department of Archaeology, Cambridge University.

Surveying and Charsadda Pakistan, courtesy of Cameron Petrie, Department of Archaeology, Cambridge University.

John Krigbaum recording human skeletal remains from Niah Cave, Sarawak.

Photograph: L. Lloyd-Smith.

*Archives des députés allemands*: [www.flickr.com/photos/hamadryades/2549161782/](http://www.flickr.com/photos/hamadryades/2549161782/)  
Harrison Excavation Archive field notebook, image courtesy of the Sarawak Museum

Rob Law printing out PhD Thesis. Photograph: L. Lloyd-Smith

Slide 4: DCC Data Lifecycle and image of *Checklist for a Data Management Plan* courtesy of the Digital Curation Centre

### *Creative Commons Licence*

The teaching materials are released under Creative Commons licence UK CC BY-NC-SA 2.0: By Attribution, Non-Commercial, Share-Alike. You are free to re-use, adapt, and build-upon the work for educational purposes. The material may not be used for commercial purposes outside of education. If the material is modified and further distributed it must be released under a similar Creative Commons licence.

