DataTrain

Open Access Post-Graduate Teaching Materials in Managing Research Data in Archaeology

Looking after Digital Research Data: Now, later, and long-term

Introduction

Looking after digital data is central to good research. We all know of horror stories of people losing or deleting their entire dissertation just weeks prior to a deadline! But even before this happens, good practice in looking after research data from the beginning to the end of a project makes work and life a lot less stressful. Defined in the widest sense, digital data includes all files created or manipulated on a computer (text, images, spreadsheets, databases, etc). With publishing and archiving of research increasingly online we all have a responsibility to ensure the long-term preservation of archaeological data, while at same time being aware of issues of sensitive data, intellectual property rights, open access, and freedom of information.

The DataTrain teaching materials have been designed to familiarise post-graduate students in good practice in looking after their research data. A central tenet is the importance of thinking about this in conjunction with the projected outputs and publication of research projects. The eight presentations, followed by group discussion and written exercises, follow the lifecycle of digital data from pre-project planning, data creation, data management, publication, long-term preservation and lastly to issues of the re-use of digital data. At the same time the course follows the career path of researchers from post-graduate research students, through post-doctoral research projects, to larger collaborative and inter-disciplinary projects.

The teaching material is targeted at co-ordinators of Core Research Skills courses for first year post-graduate research students in archaeology. The material is open access and you are invited to re-use and amend the content as best suits the requirements of your university department. The complete course is designed to run either as a four hour half-day workshop, or 2 x 2 hour classes. Alternatively, individual modules can be slotted into existing data management and core research skills teaching.

Aims & Objectives

The aim of this course is to equip first year post-graduate students with essential skills in looking after their research data for their PhD. By teaching data management as part of core post-graduate research skills training, habits of good practice will be established at an early stage of a researcher's career which they will take forward through the rest of their work.

Beyond and outside of academia, a demonstrable proficiency in data management is a highly valued cross-disciplinary and transferrable skill.

By the end of the course the students will:

- Understand the bigger issues relating to the use and archiving of digital data;
- Know of the data management requirements and resources of their university;
- Be aware of national online resources in research data management;
- Be prepared for data management in the real world after completing their PhD.

On a practical level, each student will:

- Be proficient in looking after their own digital research data;
- Have a Post-Graduate Data Management Plan in place;
- Have started to make plans for archiving the digital data from their PhD.

Teaching Materials - Design and suggestions for re-use
(L. Lloyd-Smith, July 2011)
There are eight modules each comprising a powerpoint presentation with accompanying notes. The majority of presentations end with group discussion and/or written exercises.

Modules:

- 1. [Creating and managing research data in archaeology: an overview](#)
- 2. [Data lifecycles and management plans](#)
- 3. [Working with digital data](#)
- 4. [Rights and digital data](#)
- 5. [E-Theses and supplementary digital data](#)
- 6. [Archiving digital data](#)
- 7. [Post-Graduate data management plans](#)
- 8. [Project and professional data: data management on post-doctoral research projects and beyond](#)

The first draft of the course material was written between November 2010 and February 2011. A pilot course was run as 4 x 2 hour sessions in March 2011 for post-graduate students in the Department of Archaeology, University of Cambridge. The teaching material was put together from a student's perspective and while students are encouraged to look through the course material on their own, the presentations, accompanying notes, and exercises have been designed to be led by a course leader. The modules provide a structured forum for discussions where the students learn good practice in data management by talking to each other and learning from their peers. As somebody who had obtained a PhD two years earlier, I designed the course around the two questions:

- What do students need to know about data management to get through their PhD?
- If I were starting a PhD now, what would be really useful to know about in terms of data management and archiving?

The slides and notes have been kept as simple and as straight forward as possible. They are not meant to be exhaustive in the information they contain. Rather, they provide an overview of the general issues regarding data management. The course leaders are encouraged to embellish the modules with images, local case studies and examples of good and bad data management practice from their own research and/or own department that they and the students, are familiar with. Recently completed post-graduate students might like to provide case study examples of how they managed their research data, similar to those provided as examples in Module 7.

When trialing Module 8 (Data Management on Post-doctoral and Larger Research Projects), examples of three projects based at the McDonald Institute for Archaeological Research, University of Cambridge, were used as case studies for a group discussion exercise. Research Associates from each project were present to take part in the discussions and answer questions; an exercise which they all also found useful. The project summaries and discussion questions are provided here with Module 8 as a possible template for replacing with local case studies from the department where the course is taking place.

The teaching modules were run as a trial course in March 2011, as part of a post-graduate course in Digital Skills for Dissertation and Publications, Department of Archaeology, University of Cambridge. The data management course comprised 4 x 2 hour sessions:

1. Creating and Managing Data
Defining post-graduate research data

2. Working with Digital Data
File structure, naming, and formats
E-theses and supplementary digital data
Post-Graduate Data Management Plans

3. Project and Professional Data
Data management for larger research projects

4. Archiving and Re-using Data
Depositing digital data
Intellectual Property Rights and research data

When turning these into an open educational resource the modules were broken down into eight separate sessions and the slides stripped down and condensed. The eight sessions are designed to form a complete course, with exercises leading on from one another; the 'answers' the students write down in each being referred to in later exercises, and ultimately building up to provide the necessary base to start writing a Post-Graduate Research Project Data Management Plan. It is highly recommended that the Post-Graduate Data Management Plan become an integral part of doctoral students' first year report. Departments may also like to require an updated Data Management Plan as part of doctoral students' second year reports if these are written.

Each module has been designed to take approximately 30 minutes to complete. Six of the eight presentations have between 10 and 16 slides (including front title and end

acknowledgement slides). The two longer modules are Module 3: Working with Digital Data; and Module 8: Project and Professional Data.

Module 3 (Working with Digital Data) has 38 slides many of which contain a lot of information on different file types and formats. This information has been summarised from the Archaeology Data Service's Guides to Good Practice, and content most relevant to post-graduate students is presented in a straight forward way. Rather that spending an hour presenting Module 3 in detail (and boring the students to death), it is suggested that the slides be presented as a 'lightening tour' of the practical issues of working with digital data. The slides can then be made available for future reference by the students as a handout.

Module 8 (Project and Professional Data) provides an introduction into data management at a higher level of research, including writing AHRC Technical Appendices. While this can be run as a stand alone session, given that this is the desired career path of many doctoral students, and the fact that many doctoral students carry out their research as part of larger projects, the aim of the module is to round off the post-graduate course by looking forward beyond the submission of a PhD Thesis.

The eight modules are designed to form a complete course, either as a four hour, half-day workshop or 2 x 2 hour classes (Modules 1-4 and 5-8). Alternatively, individual modules can be slotted into existing data management and core research skills teaching.

Creative Commons Licence
The teaching materials are released under Creative Common licence UK CC BY-NC-SA 2.0: By Attribution, Non-Commercial, and Share-Alike. You are free to re-use, adapt, and build upon the material 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.

Project Team
The DataTrain Archaeology teaching material was created by Lindsay Lloyd-Smith. The DataTrain Project Manager was Elin Stangeland (Cambridge University Library). Project advisors were: Stuart Jeffrey (Archaeology Data Service), Sian Lazar (Department of Anthropology, University of Cambridge), Irene Peano (DataTrain Project Officer: Social Anthropology), Cameron Petrie (Department of Archaeology, University of Cambridge), and Grant Young (Cambridge University Library).

Contact
To monitor the up-take and re-use of the DataTrain material, if you make use of any of the modules or if you have any queries or comments regarding the materials, please contact:

Elin Stangeland,
Cambridge University Library
West Road, Cambridge, CB3 9DR
Tel: 01223 333 130
E-mail