DATA MANAGEMENT FOR SOCIAL ANTHROPOLOGY - OVERVIEW

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. 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 being online, we all have a responsibility to ensure the long-term preservation of our research 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.

This teaching package is focussed on data management for Social Anthropology.

OVERVIEW OF COURSE MATERIALS

The following materials are available for each course:

- Notes and PowerPoint presentations for a series of three course modules. Each is intended to run for 1.5 hours (minimum), and designed to encourage active participation and knowledge sharing from students. The duration might therefore change depending on the expected level of participation, as well as on instructors' preferences. The notes are detailed, providing the course instructors with background information and examples (though these should be supplemented or replaced with examples from the instructor's own department and experience in order to make the course as relevant as possible), while the PowerPoint slides are designed to act as prompts for key concepts and points of discussion.

- A survey model, to be distributed in advance of each session (either once, should more than one session be run, or repeatedly, with ad hoc modifications made by the instructor for each session), to find out about students' skills, requirements and interests in relation to data management. This will allow for better participation, and students may be asked to volunteer their practical skills to the rest of the class.

- A list of useful software; this is a helpful addition to each of the three main modules.

- A list of references and web-based resources that can be handed out to participants for further reference.

Course materials are available for download from the Downloadable resources link on the left-hand menu.

BASIC MODULE

This is aimed at pre-fieldwork PhD students (and possibly at masters students, whose course requirements include fieldwork). It touches upon the following topics:

- The process of fieldwork, the kinds of data collected, and the methods for their collection. This can include a participatory discussion in which participants expose their research plans and anticipated data collection and management strategies and issues, modelled on that held in the workshop.

- The organisation of such data:

  o basic information on file management (formats, structure, naming, version control, documentation, selection - special reference to audio-visual material, where these issues are most salient)

  o some practical demonstration of software tools (databases, reference managers, and other organisation tools), and practical experience of workshop leader or other participants. This section will build on participants' own experience as much as on the trainer's own (prior to the workshop it is advised that a questionnaire be distributed or a discussion held to find out what the requirements are, and which expertise participants can contribute).

  o back-up techniques

Course contents:

1. What is data management, and why is it relevant?
   - Funding and governance
   - Personal organization and skills
2. What is data?
3. Data creation, capture and organisation Introduction
   - Some tips for the conversion of data into different formats
   - File formats
   - File naming
   - File structure
   - E-mail
   - Remote access
4. Back-ups

ADVANCED MODULE

This course builds on from the basic course. This is also aimed at post-graduate, pre-fieldwork research students, but lends itself to participation by larger audiences (writing-up students, post-doctoral researchers, senior researchers).
The following topics will be covered:

- Metadata, ontology, and other data organisation, with examples from repositories
- Ethics, IPR, FoI and Data Protection
- Sharing and dissemination tools, techniques, and experiences.

Course contents:

1. Documenting your data
   - Paper archives and indexes
   - Digital archives
   - Brainstorm
   - Techniques and software
   - Categories and ontology
   - Sharing
2. Ethical and legal issues Risks and issues in data dissemination
   - Some tips and techniques to make data safer
   - Data Protection Act 1998
   - Freedom of Information Act 2000
   - Intellectual Property and copyright

WRITING UP MODULE

This course is aimed at PhD students who have already conducted fieldwork, and are engaged in the process of writing their dissertation and thinking about future career plans; and at early-career researchers. However, like the other modules it lends itself to participation by larger audiences (established scholars and interested pre-fieldwork graduate students). Like the advanced course, it presupposes some basic data management skills (cf. basic module). It contains information on issues to do with long-term digital preservation and data sharing; funding bodies' requirements in terms of data management and sharing; digital repositories; ethics, freedom of information, intellectual property and data protection.

Course contents:

1. Digital curation and preservation - some principles and tips
   - General principles
   - Sharing
   - Digital Repositories
   - E-theses and publication
   - Funding, data management and sharing
2. Ethical and legal issues Risks and issues in data dissemination
   - Some tips and techniques to make data safer
   - Data Protection Act 1998
   - Freedom of Information Act 2000
   - Intellectual Property and copyright

The material presented here repeats some of the content of the advanced module, most notably the section on ethical and legal issues, the sub-section on sharing, and parts of the sections on digital repositories; and on e-theses and publication.

ADVICE FOR RUNNING A 1-DAY COURSE

Whilst meant to be run separately, and, in the case of the third module, for different audiences, the courses can be combined to form a one-day interactive workshop. Should this option be chosen by the instructor, the following format is recommended (please refer to the individual modules' content, and note that the total duration will not reach the sum of the three separate modules' duration):

1. What is data management, and why is it relevant?
   - Funding and governance
   - Personal organization and skills
2. What is data?
3. Data creation, capture and organisation
   - Introduction
   - Some tips for the conversion of data into different formats
   - File formats
   - File naming
   - File structure
   - Email
   - Remote access
   - Categories and ontology
   - Sharing (incl. Digital Repositories; e-theses and publication)
   - Software - some practical examples
4. Back-ups
5. Ethical and legal issues Risks and issues in data dissemination
   - Some tips and techniques to make data safer
   - Data Protection Act 1998
   - Freedom of Information Act 2000
   - Intellectual Property and copyright
6. Working with data at different levels of research: Funding, data management and sharing