

Data management for writing-up social anthropology PhDs and early-career researchers – an interactive workshop

Introduction

This is meant to run as a 1.5 hours (minimum) module, aimed mainly at writing-up PhD students and early-career researchers but lending itself to participation by larger audiences (established scholars and interested pre-fieldwork graduate students). It is designed as an interactive workshop, in which participants are encouraged to intervene and contribute their knowledge, and it presupposes basic data management skills (cf. basic module). The ‘writing-up’ module builds on the basic ‘data management for social anthropology’ course, which might be run separately or attached to the advanced and writing-up courses. Should the module be run on its own, rather than as a follow-up to the basic module for the same participants, a survey could be distributed in advance of the session, to find out about students’ skills, requirements and interests in relation to data management (see separate document for a possible survey template). This will allow for better participation, and students may be asked to volunteer their practical skills to the rest of the class. A reading list is also available, with key readings starred so that students might go to the workshop prepared for discussion on selected topics.

The duration might therefore change depending on the expected level of participation. The material presented here repeats some of the content of the advanced module (section on ethical and legal issues, sub-section on sharing), so if instructors wish to collate all three modules, parts of them, or two modules, they should note that the total duration will not reach the sum of the three separate modules’ duration.

This module contains information on issues relating to:

- long-term digital preservation and data sharing;
- digital repositories;
- ethics, freedom of information, intellectual property and data protection;
- funding bodies’ requirements in terms of data management and sharing, data management plans.

Attached to all three is a list of web-based resources that can be handed out to participants for further reference (see separate file).

All references cited in what follows have been included in the handout that accompanies these notes and the related slides. It is advised that at least some of the webpages are shown to students during the workshop, but please note the importance of verifying that links still work and, more generally, to ensure that the information provided is up to date! Finally, make sure that you substitute all institution-specific references as appropriate.

[Note: italic text between square brackets is intended as notes for the instructors]

Course Contents [slide 2]

| | |
|---|----|
| 1. Digital curation and preservation – some principles and tips | 3 |
| i. General principles | 4 |
| ii. Sharing..... | 5 |
| iii. Digital Repositories | 6 |
| iv. e-theses and publication..... | 8 |
| 2. Ethical and legal issues | 11 |
| i. Risks and issues in data dissemination | 11 |
| ii. Some tips and techniques to make data safer | 12 |
| iii. Data Protection Act 1998 | 13 |
| iv. Freedom of Information Act 2000 | 15 |
| v. Intellectual Property and copyright..... | 15 |
| 3. Funding, data management and sharing..... | 19 |

[slide 3] N.B.: this is a participatory exercise! Please feel free to interrupt, contribute, ask questions at any point...!

1. Digital curation and preservation – some principles and tips [slide 4]

[slide 5] *[Participants introduce themselves, their research and career stage, briefly indicating how they manage their data, their expectations for the workshop and specific issues they would want to discuss]*

[slide 6] Data management becomes increasingly important as you move on in your career, when time becomes scarcer and you are engrossed in multiple tasks. Today we examine some issues with so-called ‘data curation,’ i.e. the selection, preservation, maintenance, collection and archiving of data. We focus mainly on digital material, since it is increasingly predominant, as well as being the focus of data management governance.

Within the academic environment, it is crucial especially for larger, more collaborative, inter-disciplinary and cross-institutional research projects, which are becoming increasingly common given funding trends *[here examples may be given specific to instructor’s knowledge and institutional environment]*. [slide 7] In recent years, there has been a shift in patterns of funding and in funding bodies’ requirements. Funding bodies such as the AHRC and the ESRC now require that data produced through their funding be made available to the public, and therefore introduced the concept of a ‘data management plan.’ For these purposes, ‘data’ is defined as anything ‘machine-readable’ – so handwritten notes, tapes, printed images and other objects are not per se included, though of course they might be digitized. At present, this requirement only applies at post-doctoral level and above, although it may well change in the near future to include PhDs. Either way, the idea is that researchers should be trained early on in the art of managing their data for re-use, and students are ‘strongly encouraged’¹ to deposit their data and make it public. We will examine some funding bodies’ requirements in more detail later.

Regardless of funding and governance issues, as a researcher you might want to revisit fieldnotes collected in earlier times – another reason to archive your material properly, to be able to retrieve information. Whilst in the short term you will be able to rely on memory, it will fade as time goes by.

Or you might think others will be interested in revisiting this data – a lot of classical anthropological studies have served as the basis for later research and analysis, and archival material has constituted an essential part of that process. [slide 8] See for example:

- Annette Weiner’s re-study on Trobriand Island exchange practices² which of course relied on Malinowski’s own work (including the diaries, correspondence, and pictures, held at the LSE³). Weiner’s own archive (1970-1997) is held at

¹ cf. ESRC postgraduate funding guide, <http://www.esrc.ac.uk/funding-and-guidance/guidance/postgraduates/PFG.aspx>. This is likely to be reviewed for studentships commencing in October 2011, so instructors should monitor any changes in data management requirements. For the general ESRC policy on data, cf. http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf.

² Weiner, A. 1976 *Women of Value, Men of Renown: New Perspectives in Trobriand Exchange*. Austin: University of Texas Press; 1986 *Inalienable Wealth*. *American Ethnologist* 12(2):178-183; 1988 *The Trobrianders of Papua New Guinea*. New York: Holt, Rinehart & Winston.

³ http://www2.lse.ac.uk/library/archive/holdings/malinowski_bronislaw.aspx

NYU,⁴ and contains administrative memoranda, correspondence, reports, field notebooks, publications, unpublished writings (reviews, course materials and dissertations), newspaper clippings and papers produced since Weiner's graduate years.

- SOAS-led research project on Indian village life, a 'restudy' of three villages in the Indian states of Gujarat, Madhya Pradesh and Orissa that were the subject of now-classic 'village ethnographies' in the 1950s. They aim to survey living conditions in the villages as well as villagers' attitudes towards social change, and then compare their results with data from the 1950s to see how the post-colonial Indian village has changed socially, economically and politically.⁵

[slide 9] Also, in view of publishing a monograph out of your thesis, you might want to think about documentation and metadata, not only for field material but also for bibliographic references and secondary sources of various kinds. Or you might deposit your thesis in an online repository.

Finally, data management is also an asset for non-academic careers, a 'transferable skill.' And regardless of whether or not you intend to pursue an academic career, you might want to consider how to store and disseminate your material for different purposes and through different mediums in the future.

i. General principles [slide 10]

Depending on which stage of your career you are in, you might be thinking about archiving material you already have, or planning a new research project from scratch. In this course we cover all aspects of what is called the 'data life cycle' - project planning; data collection; data analysis; data distribution and archiving; data discovery and re-use; leading back to data re-analysis etc. Of course, the earlier in a research project issues of data management are taken into account, the easier it will be to organize and store data with a long-term vision.

When beginning a new project, you need to think of all the general types of analogue and digital data – e.g. digital images, audio recordings, text files, emails, spreadsheets, but also handwritten notes, artefacts and other objects - and about the specific file formats that will be used. Some formats are particularly recommended, especially for long-term preservation.⁶ *[show file formats table]* **[slide 11]**

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

For all types of digital data, it is advisable to estimate the expected volume of it so as to have enough storage space provisions available. This is also relevant for the production of data management plans, to which we will return later.

Furthermore, some thoughts should be given to the structure and naming of project folders and files, the formats and software that you will use.

⁴ <http://dlib.nyu.edu/findingaids/html/archives/weiner.html>

⁵ <http://www.soas.ac.uk/news/newsitem67996.html> retrieved 6 July 2011.

⁶ <http://data-archive.ac.uk/create-manage/format/formats-table>

Another issue to consider is how and where the different types of data will be documented – i.e. how data are created or digitised, what they mean, what their content and structure is, and whether/how they might have been modified. You might want to follow standardised terminology to describe the data, and provide contextual information for the data, in the file name or in a separate file, e.g. a spreadsheet or text file.

ii. Sharing

Whilst data documentation is useful for personal access to data, it is even more relevant if data is to be shared or made public - so that other users can understand and re-use it correctly.

[slide 12] *[Participants are asked about the sharing of their data: have they ever shared/do they plan to share data, with whom, for what purposes, how? What precautions did they take to protect privacy?]*

[slide 13] You might want to share fieldwork data with

- your supervisor/mentor;
- peers and colleagues;
- other academics and the broader research community;
- museums;
- research participants;
- wider audiences.

Of course, there are issues of privacy and intellectual property, which we will deal with in more detail later. Depending on who you are sharing with, different techniques may apply and different issues might arise.

[slide 14] As a general rule, if you know you are going to share large numbers of files, and especially if more than one person is likely to work on them, it is good to agree on certain rules and conventions as to the modality of sharing and the naming of files, and to find the most appropriate, effective, and supple way of sharing.

[slide 15] Tools useful for sharing and working collaboratively:

- Institutional networked storage

PWF:⁷ public workstation facility, giving users (limited) storage space, can be accessed remotely by multiple users *[this is institution-specific - replace with appropriate, specific content, where available, and give practical examples/tours]*

- Virtual learning or research environments

CamTools:⁸ online teaching and resource storage facility, managed by an administrator, with various levels of permission for different users. Has announcement, wiki and chat functions, among others *[this is institution-specific]*

⁷ <http://www.cam.ac.uk/cs/pwf/>

⁸ <https://camtools.cam.ac.uk/access/content/public/help.html>

- *replace with appropriate content, where available, and give practical examples/tours]*

- Dropbox:⁹ up to two gigabytes of online file space free (and up to 300 gigabytes for a fee), increased periodically; ability to share folders with other users; ability to synchronise versions of your files between different devices (e.g. your laptop, your desktop, the online space); automatic backup.
- GoogleDocs:¹⁰ especially helpful to edit documents collaboratively, to avoid using email attachments – but requires that you work online.
- Google+:¹¹ designed especially for online sharing, allows to create different circles
- Academic web networks: see especially the Open Anthropology Cooperative (OAC)¹² and Academia.edu.¹³
- Blogs:¹⁴ also good for dissemination. Some famous anthropology blogs include Savage Minds¹⁵ and Anthropologists for Justice and Peace.¹⁶
- Wikis (good for dissemination too)¹⁷
- Other project websites¹⁸
- Posting things on CDs/DVDs might be a good idea for infrequent sharing of large amounts of data. Beware of security issues, which can be sidestepped by encryption (more later); and of decay/damage.
- Whilst not ideal (attachments clog inboxes and may cause problems when downloading them), email is convenient. If you do use it, make sure one person is responsible for keeping track of the latest version of a file; beware of security privacy and data mining issues, especially salient with the main corporate email providers such as Gmail, Microsoft, Yahoo; delete messages from your mailbox.
- Digital repositories and data centres

iii. Digital Repositories

Digital repositories maintain your digital files and ensure that they remain usable over time. Repositories also provide online access to papers and data for the research

⁹ <https://www.dropbox.com/>

¹⁰ <http://docs.google.com/>

¹¹ <https://plus.google.com>

¹² <http://openanthcoop.ning.com/>

¹³ <http://www.academia.edu/>

¹⁴ <http://en.wikipedia.org/wiki/Blog>; for blog platform comparison, see <http://www.weblogmatrix.org/>

¹⁵ <http://savageminds.org/>

¹⁶ <http://anthrojustpeace.blogspot.com/>

¹⁷ http://en.wikipedia.org/wiki/Wiki_software;

http://en.wikipedia.org/wiki/Comparison_of_wiki_software

¹⁸ E.g. <http://www.oralliterature.org>; <http://www.digitalhimalaya.com>

community. They can serve as a method of publishing files and data, making them more easily citable as well as accessible. Some are organised by subject (e.g. archaeological data, historical analyses, chemical data, etc.) while others are organised by institution (e.g. materials from members of a university, usually focused on publications and theses rather than data). They range from small specialist collections to national or international services.

Repositories also provide support for documenting and annotating (metadata) and many provide additional services such as advice and assistance with data management, formats, security, and intellectual property rights concerns. Funding bodies' policies increasingly require the depositing of publications (and sometimes of data itself) in repositories such as the Economic and Social Data Service (ESDS), [slide 16]¹⁹ a national data archiving and dissemination service which came into operation in January 2003. The service is a jointly-funded initiative sponsored by the ESRC and JISC, and hosts, among others, the data produced by ESRC-funded research projects. Using the keywords 'social anthropology,' 'cultural anthropology,' 'ethnography,' their catalogue returns 64 records. The ESRC website also hosts a Research Catalogue,²⁰ containing details of ESRC-funded research projects, including details of over 100,000 research outputs (such as articles, books, papers and journal articles). There are also details of the outcomes of the projects, and the impacts that the research has had on the economy, society and individuals.

Some Higher Education Institutions also have their own institutional repositories. Many institutional repositories initially focussed on research outputs and some still limit their collections to this type of content. Others have started to widen the original remit to include learning and teaching materials. Whilst institutional virtual learning environments have, to some extent, acted as stores for learning and teaching materials, they tend not to support the search and retrieval functions required for a repository.²¹

DSpace@Cambridge, the institutional repository of the University of Cambridge, yields 1009 hits for a search with keyword 'anthropology' (some of them 'dark' items). It is currently hosting the scholarly works of 17 authors and the PhD theses of 7 students in social anthropology; 83 occasional papers in sociology and anthropology; and several anthropology lectures.²² *[should the institution at which the workshop is led have an equivalent repository or similar service provisions, these should be flagged.]*

A typical Institutional Digital Repository is usually not set up to handle or look after sensitive personal data in deposited material (more on data protection later). While some repositories may take simple self-contained HTML files, very few are set up to archive complex, interactive websites or objects.

If you think you are going to deposit all or part of your data, you might have to negotiate its sharing with your research subjects (seeking consent *post hoc* is a real nightmare usually). When planning to deposit research data it is useful to ask who

¹⁹ <http://www.esds.ac.uk/>

²⁰ <http://www.esrc.ac.uk/impacts-and-findings/research-catalogue/index.aspx>

²¹ <http://www.jiscinfonet.ac.uk/infokits/repositories/types>

²² <http://www.dspace.cam.ac.uk/simple-search?query=anthropology&x=0&y=0>.

might be interested in re-using the data you have produced. This might help in coming up with the best plan for disseminating the results of your research and the supporting digital data.

Ahead of deposit, digital repositories usually provide comprehensive guidance on issues such as formats, file structure, documentation, legal and ethical issues, encryption and migration (i.e. the transfer of digital resources from one hardware/media and software/file format generation to the next).

Following receipt of the data collection a deposit review is usually conducted. In the case of the ESDS, checks are carried out to ensure that the deposit is complete; the data collection has no ethical, legal or rights issues which might prevent sharing; the condition and format of the dataset are suitable for secondary use and long-term preservation; and documentation is sufficient to enable secondary use of data. After the initial review, the data collection is examined in detail and decisions are made regarding file preservation and dissemination formats, composition of documentation and levels of validation.

[slide 17] For a comprehensive list of research data repositories, see <http://www.datacite.org/repolist>; http://oad.simmons.edu/oadwiki/Data_repositories

List of open-access digital repositories: <http://www.opendoar.org/>

Other examples of digital repositories:

- UK Data Archive:²³ it curates the largest collection of digital data in the social sciences and humanities in the United Kingdom, hosting several thousand datasets. It was established in 1967 through the Social Science Research Council, which has provided the long-term commitment of funds. It acquires data from the academic, public, and commercial sectors, and manages the Economic and Social Data Service. They also host a number of data services such as the Census portal and the History Data Service, and are engaged in a number of data management initiatives.
- EThOS (online theses)²⁴ – although it is moving towards a database aggregating resources held in institutional repositories by providing a search interface across the UK (more below).

iv. e-theses and publication **[slide 18]**

It is increasingly common for higher-education institutions to provide facilities to deposit electronic copies of PhD theses in digital repositories. At Cambridge, this is done through DSpace@Cambridge²⁵ *[replace with institution-specific information where applicable]*, and the British Library also provides an online repository service for theses.²⁶ EThOS is a searchable gateway to new, non-embargoed e-theses from

²³ <http://www.data-archive.ac.uk/>

²⁴ <http://ethos.bl.uk/Home.do;jsessionid=4E4329AACA0E77CF0F0C390C3A8648C5>

²⁵ <http://www.dspace.cam.ac.uk/>

²⁶ <http://ethos.bl.uk/Home.do;jsessionid=4E4329AACA0E77CF0F0C390C3A8648C5>

contributing universities, a list of which is available on their website. They also scan existing paper theses manuscripts from some institutions on request. In some institutions or departments the requirement to deposit an online copy of one's PhD is mandatory.

If the submission of an e-thesis is optional there are a number of reasons for and against putting your PhD online. Reasons for depositing an e-thesis might be:

- Making the findings of your research available to all.
- Raising your profile amongst your research community.
- Long-term archiving of your PhD with a persistent internet address (URL).

At the same time, there are valid reasons for not putting your thesis online:

- Publication plans for the thesis: your thesis publication plans might be hindered by making the thesis available on-line before hand. Some publishers allow on-line release of theses, others do not – so before you make a decision it is advisable you find a publisher and check with them. However, although an approved PhD is normally a piece of work of publishable standard, examiners might provide recommendations for amendments to be made prior to submitting the whole or parts of it for publication. In this sense, your monograph might turn out to be quite different from your PhD and you may therefore be able to deposit the thesis online without any copyright issues.
- The thesis contains sensitive data: before making any of it public you need to make sure you have taken all the necessary precautions – anonymisation is advisable, but it might not always be possible to conceal details about people and places. Another way around it could be to put an embargo on the thesis, which will make it public only after a number of years.
- The thesis contains a significant quantity of copyrighted material. In all instances where you are using data and other material protected by copyright, permission has to be granted by the relevant authorities or individuals. There are several options for theses which include third-party copyright material.²⁷ Depending on the quantity of such material, different measures may be adopted:
 - Obtain permission to use copyright material. Permission letter templates are available online, for example from the [jisclegal.ac.uk](http://www.jisclegal.ac.uk) website.
 - If only a few items in the thesis are concerned, either leave them out of the thesis or withdraw copyright materials from the digital version. A note should be inserted to explain why it was removed.
 - If the thesis includes a substantial amount of copyright material and/or permission is difficult to obtain, then it may be best not to make the thesis available online.
 - Put an embargo on the thesis.

²⁷ Copyright guidance written by the Cambridge legal services can be found here: http://www.lib.cam.ac.uk/repository/theses/theses_permissions.html [please replace or supplement with institution-specific information]

Your supervisor is usually the first reference point to discuss these issues. We will come back to copyright and other issues to do with sensitive data and legal requirements later.

In addition to your thesis, you might consider archiving parts of your data in a repository – for example information included in appendices in the form of tables, if that is part of your corpus. For the publication of data, several options are available apart from institution-specific repositories – e.g. the ESDS.

2. Ethical and legal issues [slides 19 and 20]

The dissemination of data comes with thorny ethical issues, and potentially relates to legal ones as well. A lot of the data produced through fieldwork is in one way or the other ‘sensitive,’ if only because it contains personal information concerning research subjects and the researcher her/himself. Therefore, when creating and storing data one needs to evaluate possible risks and consider taking appropriate measures. Ask yourself who might have access to your data (on your computer, USB stick, online cloud, notebooks...).

i. Risks and issues in data dissemination [slide 21]

Risks [slide 22]:

- Online storage: clouds and emails are handy, but potentially unsafe. Security for established online backup and sharing services is decent, but not guaranteed, and some of the intellectual property rights agreements for the sites are a bit vague; you should encrypt your files if they contain particularly sensitive data (more later). There is always the possibility that your online service will go out of business, leaving you without your important files. Finally, beware of the Data Protection Act (see below).
- We also mentioned the issue of sharing and consent: if you plan to share data, you should inform research participants, ideally as early as possible.
- Crossing borders: under some legislation such as that of the US (of the ‘sovereign exception’ kind) electronic equipment can be randomly seized and searched at international borders, in view of ‘terrorist threats.’²⁸ More widely applicable, international agreements that purportedly seek to combat ‘counterfeiting’ are being negotiated, which would have similar consequences.²⁹ Depending on which borders you cross, you might need to take this into consideration.

Different ethical issues push for sharing or withdrawing data. Here is what the Association of Social Anthropologists of the UK and Commonwealth (ASA) ethical guidelines say on the issue [slide 23].³⁰

(3) Sharing research materials: Anthropologists should give consideration to ways in which research data and findings can be shared with colleagues and with research participants:

(a) Research findings, publications and, where feasible, data should be made available in the country where the research took place. If necessary, it should be translated into the national or local language. Researchers should be alert, though, to the harm to research participants, collaborators and local colleagues that might

²⁸ http://en.wikipedia.org/wiki/Border_search_exception;

²⁹ http://en.wikipedia.org/wiki/Anti-Counterfeiting_Trade_Agreement#Border_searches

³⁰ <http://www.theasa.org/ethics/guidelines.shtml>. Please note that at the time of writing this module, the guidelines are being updated, so refer to the website for the latest version. Also see the American Anthropological Association guidelines and resources on ethics at <http://www.aaanet.org/committees/ethics/ethics.htm>

arise from total or even partial disclosure of raw or processed data or from revelations of their involvement in the research project;

(b) Where the sharing with colleagues of raw, or even processed, data or their (voluntary or obligatory) deposition in data archives or libraries is envisaged, care should be taken not to breach privacy and guarantees of confidentiality and anonymity, and appropriate safeguards should be devised.

(4) Collaborative and team research: In some cases anthropologists will need to collaborate with researchers in other disciplines, as well as with research and field assistants, clerical staff, students etcetera. In such cases they should make clear their own ethical and professional obligations and similarly take account of the ethical principles of their collaborators. Care should be taken to clarify roles, rights and obligations of team members in relation to matters such as the division of labour, responsibilities, access to and rights in data and fieldnotes, publication, co-authorship, professional liability, etcetera.

So, to share or not to share? Ultimately, this is a context-specific choice which might have to be re-negotiated in time.

ii. Some tips and techniques to make data safer [slide 24]

- For important documents and to enter your hard-drives, restrict access by setting up passwords and access permissions (e.g. no access, read only, read and write, administrator-only permission).
- Make sure you log out of websites (e.g. social networks, online banking, email...) and of your hard -drives if you aren't using them
- Use firewalls and anti-virus software (esp. for Windows)
- Always keep multiple copies, in different locations and formats
- Destroy data when needed - both hard- and e-copies (n.b.: data deleted from a hard-drive might still be retrievable, as it is from emails etc. - which may be good or bad news)
- Encryption: this is a more secure way of both storing and transferring data. Software is available to perform this, both proprietary (e.g. PGP) and open source (e.g. GnuPG).³¹ In transmission, a key and passphrase are used to digitally sign each encrypted file and thus allow the recipient to validate the sender's identity.
- Tiered consent: A good way to navigate the ethical complexities involved in ethnographic fieldwork is to negotiate a 'tiered' type of consent to participating in research, where possible. This involves asking subjects to consent to participating in the study; having notes taken of their speech; having speech recorded; having their names published; having information concerning them anonymously published; having information concerning them published with their names. Consent needn't be given to all levels - this might help you secure collaboration even when subjects may not be comfortable with use of personal information in some aspects of your research process. Of course, this is not always and in all cases feasible or possible - once again, the evaluation of specific situations is up to you.
- Anonymisation: another solution to ethical issues, especially where consent may be difficult to negotiate (as when 'data' is used from experience, from

³¹ http://en.wikipedia.org/wiki/Encryption_software

serendipitous/chance encounters, which were not formally envisaged as fieldwork in advance, for example - encounters on the bus, in bars, shops...you name it) or when participants specifically ask for it.

A few tips:

- Replace names with pseudonyms.
- Remove or alter other direct or indirect identifiers (e.g. date and place of birth and residence; age; occupation). You can often convey the sense of someone's life story, or of a situation, even if some details are altered (e.g. someone selling okra in the market might be presented as selling cooked food; a migrant who left to Germany might be said to be working in the Netherlands; age can be kept vague, e.g. a woman in her 20s). It might be difficult to conceal some locations if the events taking place or the people involved are unique in some way (so in the case of my fieldwork, Benin City is the hub of trafficking in Nigeria and there would not be much point in withdrawing the place name. But other strategies allowed to conceal the identity of individual participants).
- Be consistent in the use of pseudonyms, and keep track of them (in separate place other than the data file itself).
- When possible, negotiate in advance with subjects what needs to be concealed/alterd (such that perhaps names or other details might be omitted in interviews themselves, for example). However, you might want to consider long-term re-use of data (such that some contextual information might be useful in the future and best collected and kept, if in restricted access)
- If no other strategy works, consider restricting access to the data/thesis.
- You may not need as many alterations as you think!

For further advice, see <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation>

iii. Data Protection Act 1998 [slide 25]

The Data Protection Act 1998 was formulated in response to concerns about the amount of personal information - and the accuracy of such information - being stored, processed and passed on by organisations. It provides the legal basis for how organisations handle information relating to living people (personal data).³²

A distinction is made between 'Personal Data' and 'Sensitive Personal Data' of living individuals.

- Personal Data relates to living individuals which identifies them: name, age, sex, address, etc.
- Sensitive Personal Data is data that may incriminate a person: race, ethnic origin, political opinion, religious beliefs, physical/mental health, sexual orientation, criminal proceedings or convictions.

³² For further advice, see <http://www.jisclegal.ac.uk/LegalAreas/DataProtection.aspx>. For the full text of the Act, see <http://www.legislation.gov.uk/ukpga/1998/29/contents>. See also the more recent Data Sharing Review at <http://www.justice.gov.uk/reviews/datasharing-intro.htm>

Once a person is dead, the DPA does not affect their personal data.

A different level of security applies when dealing with Sensitive Personal Data: this could potentially apply to the contents of email and research data in some fields; students and researchers in the medical, social sciences and allied subjects are particularly urged to be aware of this requirement.

The act establishes that:

- Data may only be used for the specific purposes for which it was collected.
- Data must not be disclosed to other parties without the consent of the individual whom it is about, unless there is legislation or other overriding legitimate reason to share the information (for example, the prevention or detection of crime). It is an offence for Other Parties to obtain this personal data without authorisation. Consent forms should be made and filled and signed participants for these types of personal data to be included in project archives.
- Individuals have a right of access to the information held about them, subject to certain exceptions (for example, information held for the prevention or detection of crime).
- Personal information may be kept for no longer than is necessary and must be kept up to date.
- Personal information may not be sent outside the European Economic Area (the 15 EU member states together with Norway, Iceland and Liechtenstein) unless the individual whom it is about has consented or adequate protection is in place, for example by the use of a prescribed form of contract to govern the transmission of the data or if that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to processing of personal data.
- Subject to some exceptions for organisations that only do very simple processing, and for domestic use, all entities that process personal information must register with the Information Commissioner's Office.
- The departments of a company that are holding personal information are required to have adequate security measures in place. Those include technical measures (such as firewalls) and organisational measures (such as staff training).
- Subjects have the right to have factually incorrect information corrected (note: this does not extend to matters of opinion)

Anonymised or aggregated data is not regulated by the Act, providing the anonymisation or aggregation has not been done in a reversible way. The Act applies only to data which is held, or intended to be held, on computers ('equipment operating automatically in response to instructions given for that purpose'), or held in a 'relevant filing system'. The DPA may apply to the contents of an electronic address book, or email messages that have been backed up to a 'cloud' solution.³³

³³ For further guidance, see http://www.ico.gov.uk/for_organisations/data_protection/the_guide.aspx. Institutions should also be able to offer support in relation to this and other legal requirements.

iv. Freedom of Information Act 2000 [slide 26]

This Act was established to increase transparency in the public sector. It gives people the right to request access to recorded information held by public sector organisations or be informed whether information is held. Research data can be requested under the Freedom of Information Act, but copyright to such data stays with the original researcher.³⁴

Exceptions exist to the Act, such as:

- personal data cannot be requested
- information that is accessible by other means e.g. via a website
- information intended for future publication
- information that is subject to a confidentiality agreement, such as in a signed consent form or sensitive data held under restricted access by a data archive

Any person can request any data held by public authorities – including universities. The data does not have to have been produced by the university: it is the fact they hold the data that is important. A request must specify what data are sought.

This is a potential issue for collaborative projects where multiple copies of data are held in different institutions and countries.³⁵

v. Intellectual Property and copyright [slide 27]

“Intellectual property rights... are rights granted to creators and owners of works that are the result of human intellectual creativity”³⁶

Different forms of intellectual property are regulated by law:

- Copyright: this refers to creative works fixed in material form. Under the Copyright, Designs and Patents Act, 1988³⁷ copyright applies to:
 - original literary, dramatic, musical or artistic works
 - sound recordings, films, broadcasts or cable programmes
 - the typographical arrangement (layout) of publications
 - teaching materials and blogs

Most research outputs such as spreadsheets, publications, reports and computer programs fall under literary work and are therefore protected by copyright. Facts, however, cannot be copyrighted. Data is not covered by copyright, but the arrangement of data in a spreadsheet or database is. More generally, copyright protects the expression of an idea, not the idea itself. Copyright does not require

³⁴ For the full text of the act, see <http://www.legislation.gov.uk/ukpga/2000/36/contents>

³⁵ For further guidance on FoI and research data, see <http://www.jisc.ac.uk/publications/programmerelated/2010/foiresearchdata.aspx>

³⁶ www.jisclegal.ac.uk

³⁷ For the full text of the legislation, see <http://www.legislation.gov.uk/ukpga/1988/48/contents>

registration. In the act of creating a piece of work, writing something down, or recording an interview or song, the creator(s) of the work by default hold the right to copy the work in the future.

The ownership of copyright is not the same for all creators of work, it depends on their academic status (e.g. students or lecturers) and employment position. Students are not employees so they enjoy copyright in their own work, and some universities allow their academics and researchers the rights to their works.³⁸ Of course things can become messier when students are employed on projects and there are external funders or partners involved in projects. Different institutions have different copyright clauses in their employment contracts.

A copyright owner has the right to control the copying, adaptation, publishing, performance and broadcast of the work, and under what conditions this may be done. These conditions may involve payment of a royalty or licence fee. The owner may also give or sell some or all of the rights to others. In addition the author of a copyright work has certain “moral rights” that always remain with the author. These are the right to be identified as the author of the work, the right to object to derogatory treatment of their work and the right to object to false attribution of a work. However, these rights do not exist where copyright in a work has been originally owned by the author’s employer. The onus of responsibility lies with the user of a work to get permission, even if the rights holder is unknown or cannot be traced.

The right to use copyright material is typically obtained:

- with the permission of the copyright owner, as set out in the terms of a licence (such as the work’s terms of use), or through a licence issued by a collective licensing society which has authority to issue a licence on behalf of the copyright owner
- by seeking and obtaining permission directly from the copyright owner
- by means of an assignment (assignation in Scotland) of copyright in writing from the copyright owner.

A licence gives someone permission to do the acts which the copyright owner is entitled to authorise or prohibit without infringing copyright. This is how a great deal of material is lawfully used in the education context. In addition there are certain very specific situations where it may be permissible to make use of someone else’s copyright protected works without seeking permission from the owner. For example, it is not necessary to get permission in order to use a insubstantial (i.e. trivial) part of a copyright protected work.

Even if material is available on the Internet, permission will still be required in order to reuse the material (such as copying it, adapting it or dissemination of it by a different means or in different formats). Some websites may give information about the permissions (licence) which is granted to users, which will clarify what can and cannot be done with material.

³⁸ This is the case for Cambridge.

There are a number of exceptions in copyright law which allow limited use of copyright works without the permission of the copyright owner. In the education context relevant exceptions include:

- fair dealing for non-commercial research and private study, criticism and review
- non-exact copies of works for teaching purposes in educational establishments (such as copying material by hand)

Use of a copyright protected work without its owner's permission may be a civil infringement and/or a criminal offence depending on the circumstances. Copyright is infringed if a person does (or authorises another to do) any of the exclusive acts restricted by copyright without the permission of the owner, in relation to the whole or a substantial part of a copyright work. What amounts to a substantial part is not defined in law but it is quite likely that even a small portion of the whole work will still be a substantial part.

Copyright is essentially a private right so it is generally for the rights holder to decide what to do when his or her copyright is infringed. The infringer could be taken to court and can run the risk of having to pay compensation to the copyright owner. They could also face:

- having an injunction (interdict in Scotland) taken out against them to stop use of the material
- being ordered to surrender the copyright material to the copyright owner
- a order requiring that infringing goods be destroyed or delivered up to the copyright owner, and that any resulting profits from the infringement are paid to the copyright owner.

Where deliberate infringement of copyright is undertaken as part of a trade or business, it may be a criminal offence, punishable by an unlimited fine and up to ten years' imprisonment.

The duration of copyright may depend on whether a work is published or unpublished, whether the creator is known or unknown, and whether transitional arrangements from previous copyright legislation apply. However, in general terms, following the end of the year of the death of the creator(s), copyright lasts:

- 70 years for literary, dramatic, artistic and musical works, films and video recordings
- 50 years for sound recordings and broadcasts
- 25 years for the typographical arrangement of published editions.

Other forms of intellectual property include:

- Designs: Appearance and shape of product
- Patents: Inventions – things that make things work
- Trade marks: Signs that distinguish goods and services
- Moral rights: Right to be attributed for your work or right to object to derogatory treatment of your work.

Intellectual Property Rights can be bought, sold, rented, gifted and bequeath.
Different countries have different copyright law.

[slide 28] *[In light of what has been discussed, participants are asked to go back to issues of sharing and publishing in ethical, political and analytical terms – cf. also the reading list, especially De Nicola 2011a,c; Parry and Mauthner 2004 and Bishop 2005]*

3. Funding, data management and sharing [slide 29]

As mentioned, funding bodies are moving towards open-access policies concerning not only research outputs (publications), but the data itself. [slide 30]

The Research Council UK principles on data policy³⁹ establish that

Making research data available to users is a core part of the Research Councils' remit...Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property...To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.

RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.

To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research.

These inform the various Research Councils' policies. The Digital Curation Centre (DCC) has produced a table on funding bodies' requirements concerning data management,⁴⁰ as well as a more comprehensive overview for each.

The ESRC is the most demanding,⁴¹ stipulating that:

- Publications should be deposited at or around the time of publication.
- Data must be made available for preparation for re-use and/or archiving within three months of the end of the award.
- Researchers should submit a data management and sharing plan. At present the Council asks five questions, covering a survey of existing data, details of potential users and plans for archiving. In early 2011 a new requirement came into force, such that grant applicants will be required to submit a statement on data sharing and provide a 2-page data management and sharing plan. Nine suggested themes are provided to include in this. The ESRC will review any costs associated with implementing the data plan as an integral part of the funding decision, and based on this decision, provide appropriate funding for data management.
- The final payment of a grant may be withheld if data has not been offered for deposit to the required standard, unless a waiver has been agreed in advance.

³⁹ <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

⁴⁰ www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

⁴¹ <http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/esrc>

There is of course room for negotiation, and their stance on PhD data sharing is not yet clear. It is unlikely that they will enforce the official regulations, as this would imply a massive investment and huge costs. In any case, 'the changes introduced by the ESRC Research Data Policy in relation to data management plans will not affect existing ESRC investments and applications submitted before 19 April 2011.' (<http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>)

Other funding bodies that might be relevant to anthropology research [please check that the information is up to date at the time of the workshop]:

- Wellcome Trust: The Wellcome Trust expects all of its funded researchers to maximise the availability of research data with as few restrictions as possible. All those seeking Wellcome Trust funding should consider their approach for managing and sharing data at the research proposal stage. In cases where the proposed research is likely to generate data outputs that will hold significant value as a resource for the wider research community, applicants will be required to submit a data management and sharing plan to the Wellcome Trust prior to an award being made.⁴²
- NSF (USA): 'Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labelled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.' The data management plan can contain policies and provisions for re-use, re-distribution and the production of derivatives. The NSF policy already stipulates that grant-receiving researchers are expected to share at 'no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.'⁴³ In their guidelines, they define data as:

the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples). Research data also do not include:

- (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
- (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.⁴⁴

- The Wenner-Gren Foundation: only copies of publications resulting from the funded research are requested.
- For a comprehensive list, see <http://www.sherpa.ac.uk/juliet/index.php>

⁴² <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

⁴³ http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

⁴⁴ http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf

[Participants' suggestions on data management requirements by other funding bodies they have come across]

Most funding bodies will ask for an end-of-grant report, which is helpful to start/update the process of documenting your data.

[slide 31] In order to complete a data management plan (or 'technical appendix' as the AHRC calls it), the DCC has set up an interactive tool.⁴⁵ On their website, you will find a very detailed and somewhat complicated checklist,⁴⁶ and a template for a data management plan.⁴⁷ The Rural Economy and Land Use Programme (Relu) Data Support Service also produced a template for a data management plan,⁴⁸ and the UK Data Archive has set up a data management costing tool.⁴⁹

[slide 32] On the ESRC website you will also find some advice.⁵⁰ Questions asked in grant applications include the following:

- If the research involves data collection or acquisition, please indicate how existing datasets have been reviewed and state why currently available datasets are inadequate for this proposed research.
- Will the research proposed in this application produce new datasets?
- It is a requirement to offer data for archiving. Please include a statement on data sharing. If you believe that further data sharing is not possible, please present your argument here justifying your case.
- Who are likely to be the users (academic or non-academic) of the dataset(s)?
- Please outline costs of preparing and documenting the data for archiving to the standards required by the affiliated data support service (Economic and Social Data Service) working with the Research Councils.

According to the ESRC, a data management plan should include:

- assessment of existing data
- information on new data
- quality assurance of data
- back-up and security of data
- expected difficulties in data sharing
- copyright/Intellectual Property Right
- responsibilities

⁴⁵ <http://dmponline.dcc.ac.uk/>

⁴⁶ http://dmponline.dcc.ac.uk/system/attachments/8/original/DCC_Checklist_DMP_v3_md_sj.pdf?1300724157

⁴⁷ http://www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMP_template_v1.2_100106.doc

⁴⁸ <http://www.esds.ac.uk/create/esrc/reludmp.doc>; http://relu.data-archive.ac.uk/DMP_FR.pdf; http://relu.data-archive.ac.uk/DMP_SEIRA.pdf; see also <http://relu.data-archive.ac.uk/plan.asp>

⁴⁹ http://www.data-archive.ac.uk/media/257647/ukda_jiscdmcosting.pdf

⁵⁰ <http://www.esds.ac.uk/aandp/create/esrefaq.asp>; <http://www.esds.ac.uk/create/esrc/dataman/>

- preparation of data for sharing and archiving

[slide 33] *[Some time should be spent browsing the DCC, UKDA, Relu and ESRC tools and guidance, possibly using them with reference to concrete examples of actual research projects. This can be done collectively or in groups. Provide examples of data management plans for anthropology research projects, if possible, and ask participants to share their experience. If relevant, participants may be ask to bring their data management plans in-progress, or to draft one during the workshop]*

[slide 34] Some general points:

- Think about the ‘data life cycle’
- Questions to be answered in a data management plan may include the following:
 - What types of data will I produce?
 - From what sources?
 - How will I organise the data? (file structure/naming, formats, software)
 - At what points and how will the data be evaluated?
 - What data will be deposited and where?
 - Who will be interested in re-using the data?
 - What will the costs of managing data be?