



History DMP Project

The case for a History Data Management Plan

Richard Green and John Nicholls

November 2011



The History DMP Project

Project Director:

Chris Awre (c.awre@hull.ac.uk)

Project Manager:

Richard Green (r.green@hull.ac.uk)

Project Officer:

John Nicholls (j.nicholls@hull.ac.uk)

The History DMP Project was undertaken by Library and Learning Innovation and the Department of History at the University of Hull. It was funded by the JISC 'Managing Research Data' programme.



This material is made available under a Creative Commons Licence: Attribution-Noncommercial-Share Alike 2.0 UK: England and Wales. See: <http://creativecommons.org/licenses/by-nc-sa/2.0/uk/>

Table of Contents

1.	Project remit.....	4
2.	Fact finding.....	4
3.	Data types, formats, standards and capture methods	5
4.	Ethics and intellectual property	5
5.	Access, data sharing and re-use.....	5
6.	Short-term storage and data management.....	6
7.	Deposit and long-term preservation of data.....	6
8.	Towards a DMP	6
9.	The role of data management in Historical research	7
	Appendix 1: Core interview questions	8
	Appendix 2: Datasets discussed	9

1. Project remit

“The History DMP project will build on the established data management practices within the Department of History at the University of Hull. It will use the experience in data management demonstrated and recognised over the past few years to frame a departmental approach to data management to enhance and build on the individual activities that have been the basis of data management thus far. This departmental data management plan will look to support future research strategy and provide a coherent platform for data sustainability in future research. The work to be undertaken will develop an overall plan, and will then look to implement this using past, present and future research activities. The role of local technical provision, in the shape of the University’s Fedora-based digital repository, will be explored and enhanced in specific ways to deliver a platform that not only manages the data, but allows for its exploitation and access as well through repository, virtual learning environment, and linked data interfaces.”¹

2. Fact finding

This document represents a synthesis of findings from the project team’s initial fact finding interviews with academics in the Department of History. The individuals were chosen in such a way as to cover a range of ages, interests and data types. For convenience, the findings are here grouped using headings (sections 3 – 7 below) from the Digital Curation Centre (DCC) ‘Checklist for a Data Management Plan’.² The meetings were intended to explore two key areas:

- what are the current data management practices amongst academic staff in the department, and thus
- what are the requirements for data management in the future that should be encapsulated in a specific, departmental data management plan (DMP).

The information gathered from these interviews has also been used to inform a short report on the role of data management in supporting History research to accompany the eventual DMP (section 9).

The interviews were conducted to a common pattern based around 13 key questions: these are to be found in Appendix 1. The interviews were recorded, in order to augment interviewers’ notes, but it was made clear to participants that these recordings were not for dissemination in any form. In addition to interviews with some ten academics, a number of less structured focus groups and meetings are being held, devoted to discussions around the four types of data identified by historians:

- qualitative data
- quantitative data
- oral testimony, and
- archival data

Information from these additional meetings, which at the time of writing have not yet been completed, will be used to refine the development of a departmental data management plan.

¹ HDMP Project plan at <https://hydra.hull.ac.uk/resources/hull:4930>

² The DCC Checklist can be found via <http://www.dcc.ac.uk/resources/data-management-plans>

3. Data types, formats, standards and capture methods

Academics interviewed were found to be dealing largely with qualitative and quantitative data. Of the quantitative data, some was produced by analysis of qualitative records; other such data may have originated in numeric form. Interviewees held their data in a variety of formats ranging from wholly paper-based, through paper-based but with digitised copies, through digital versions created with now outdated database software, to formats reflecting the use of up-to-date software applications. There was little evidence of standards being used except in so far as current digital work tended to reflect formats in widespread use, mainly those deriving from the Microsoft Office suite of programs although isolated examples used Apple applications and formats. Databases were generally flat files; only a few examples of relational databases were discovered. The potential utility of parts of datasets, rather than the whole ones, was rarely mentioned by interviewees unless they were prompted.

4. Ethics and intellectual property

In discussing their datasets with us, it seemed that academic staff were well aware of issues around intellectual property and copyright. Generally, they seemed to have a clear idea of where copyright in their data lay and, if not, were clear in their minds that they should find out before committing to work on it with us.

Several of our interviewees expressed concern that, should their data be made more widely available, they should be clearly cited should it be used elsewhere.

5. Access, data sharing and re-use

During our interviews, we discovered a wide range of responses to questions about data sharing. There were clearly those who did not wish their data to be shared at all, largely so that it could not be used by others in their own field of study. At the other end of the range were staff who were more than willing to make their data available to the world – either because they were ‘finished with it’ and had produced all the outputs from it that they had intended, because they had transferred their interest to a different field of study, or because their grant funding mandated public accessibility. Between these two extremes a range of responses reflected willingness to make data available to identified groups and under controlled conditions.

Most of those interviewed recognised that should their data be made available to a wider audience it would need to be accompanied by metadata. This is a term that has a different meaning if you are a historian rather than an IT person or librarian. To a historian, metadata is about providing general context for data, effectively documentation, and the project team will need to be careful not to confuse this with the more specific IT/library use of the term.

The conversation in one interview raised the issue of a database being associated with more than one output, for example referenced as part of a PhD thesis but also in subsequent research projects. This is something that will need to be considered carefully; it points up the fact that the relationship between a digitised dataset and related research is not necessarily one-to-one.

6. Short-term storage and data management

Asked about their management of data, our colleagues probably covered the entire spectrum of behaviour from the worryingly casual to the careful and well-considered. At the one end were paper records that existed only as single copies all held in one place together; at the other were multiple digital copies held in geographically separate locations, one of which might have been a professionally maintained archive. The majority of the staff seemed to be reasonably conscientious about taking and maintaining backups of digital material.

7. Deposit and long-term preservation of data

Our interviewees were generally much less clear about their strategy for maintaining the data (or having it maintained on their behalf) in the longer term. As noted in the previous section, some colleagues had formally deposited their data with an appropriate agency but they were in the minority. We should sound a cautionary note around an individual who had deposited his dataset with a national data centre where it had been converted to a format suited to long-term preservation (tab-delimited, as it transpired) but who had difficulty accessing the copy there and could not open it once it was downloaded because Microsoft products no longer automatically open such files.

8. Towards a DMP

The colleagues that we interviewed seemed generally pleased that 'someone' was taking an interest in their data and was prepared to work with them to improve its safe keeping for the long-term and to manage its accessibility on their behalf. The idea of a checklist that might, on the one hand, help them ask the right strategic questions of themselves and, on the other, might help them identify sources of assistance to implement a data strategy appeared to be a welcome one. Some, but not all, were aware of documents like the DCC checklist but those that were agreed that it was too complex for them to cope with and, in any case, suffered from not being subject oriented. The majority of those we interviewed were willing, even keen, to help develop a departmental data management plan and, of those, many were willing to offer their data as case studies.

At the conclusion of the initial round of interviews we are satisfied that there is a need to develop a History-centric DMP for the department. Whilst this might follow the outline of the DCC checklist it will need to be clearly relevant to history research if our colleagues are to make regular use of it. The plan will need to cover management of data initially on paper as well as born-digital materials.

The plan will need to cover data at different stages in its own life cycle and at different stages in the research lifecycle. As an example of this, it should guide researchers to think about the possible long-term stewardship of their initial paper sources as well as datasets derived from their analysis which may form the basis of the published research.

We further anticipate that it should reference appropriate points of contact within the University where our academics can go for supportive advice about the management of their data throughout its lifecycle. Many were unaware who might offer advice and/or services around digitisation, hosting, data migration, long-term preservation and the like – or even that such advice and services were available.

9. The role of data management in Historical research

The one aspect of this project that is patently clear is that, in terms of data management and the role of academic historians, very little is clear! As intimated before, the historians interviewed were generally very forthcoming, eager to engage and willing to help. What also became apparent was the lack of coherent approaches in the managing of the data that is routinely gathered in the course of historical research.

Data Management, for historians, is quite often a black box which they occasionally need to access when a certain level of research has been attained. Very often, research projects require datasets or indeed databases to be developed and published. But, outside of the actual acquisition, sourcing and interpretation of relevant data, the process of data analysis (rather than the engagement with data) is not widely understood or accepted.

The implications are manifold. Clearly, at any level of data collection and collation, a flawed data management process results in flawed outputs. Often the recourse of historians is to construct convoluted and complicated systems to accommodate the difficulties they encounter. What is missing, and what may ease or even eradicate the burden, is a viable data management plan.

Specifically in the historical research context, data management is often viewed alongside historiography, and thus regarded as interesting but external to the practical aspects of the historian's work. In these terms, data management is regarded as, at best, a peripheral and optional subject rather than an integral aspect of the historical research project.

The simple truth is that, as with any prospect of data collation and dissemination, data management in historical research is vital, despite its general lack of recognition. Typically, interested voluntary enthusiasts are employed to undertake the data management tasks which they are often not qualified or capable of performing. This does not imply that these people are inefficient or incapable, but merely shows that it is often the case that convoluted, person specific systems are typically the outputs of successful projects, while just as often, despite a great deal of effort, outputs fail completely.

This highlights how the tasks of data management are regarded as secondary in terms of importance.

Appendix 1: Core interview questions

1. Which areas of your research do you feel that we may be interested in?
2. What sort of data does your research give rise to?
3. What are you doing with your data?
4. Is your data safe/secure?
5. How far along is your data in the process of finalisation?
6. What are the current prospects for your data?
7. What is being done about this?
8. What do you see as the intention for your data? (Public access/Closed access/Secure/Curateable/Other)
9. What is the life stage of your data? (Initial/Raw data/Project driven/Complete/Evolving/etc)
10. What is the format, appearance, structure of your data?
11. Is the data currently available? If yes, how?
12. Is your data currently being housed in a repository? (History data centre/University/Other)
 - a. If yes, how appropriate, useful, effective do you find this?
 - b. If no, do you want/need to have your data hosted and/or curated?
13. If appropriate, would you be interested in having your data used in a case study? This would involve making the data available, formatting, various options for dissemination and/or secure storage.

Appendix 2: Datasets discussed

This appendix is included to give the reader a broad overview, without detailed explanation, of the types of data and datasets that were discussed during the interviews that took place.

Administrative database for the University's fine art collection

Biographical detail of officers in the German armed forces in the 17th and 18th centuries

Transatlantic slave database

Slave labour during the holocaust

History of marine animal populations (HMAP) including Incofish data

Statements of navigation (1871-1913) (which informed a book)³

England's sea fisheries (economic and environmental aspects)

British privateering in the 18th century

Medieval ships in the 14th century

Deserted medieval villages in England

UK economic statistics 1066 – present

A 'living' database related to inhabitants of East Yorkshire

Public officials (sheriffs) 1280-1410

Privateering 1812-1828

US insurance industry 1850-1920

Basil Greenhill photographic collection and associated metadata

³ Starkey, David J. (ed.) (1999) *Shipping Movements In The Ports Of The United Kingdom, 1871-1913*, University of Exeter Press