Project Identifier: History DMP
Version:
Contact: Richard Green r.green@hull.ac.uk
Date: April 2012

JISC Final Report

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	History DMP		
Project Hashtag	#historydmp		
Start Date	01-10-2011	**End Date**	30-04-2012
Lead Institution	The University of Hull		
Project Director	Chris Awre, Head of Information Management, Library and Learning Innovation		
Project Manager	Richard Green		
Contact email	r.green@hull.ac.uk		
Partner Institutions	Acuity Unlimited, MediaShelf LLC		
Project Web URL	historydmp.wordpress.com ; www.hull.ac.uk/historydmp		
Programme Name	*Managing Research Data*		
Programme Manager	Simon Hodson		

Document Information			
Author(s)	Chris Awre, Martin Dow, Richard Green, John Nicholls		
Project Role(s)	Director, Contractor, Project Manager, Project Officer respectively		
Date	30-04-2012	**Filename**	HDMP-FinalReport-v10
URL	https://hydra.hull.ac.uk/resources/hull:5422		
Access	This report is for general dissemination		

Document History		
Version	**Date**	**Comments**
1.0	2012-05-02	For submission to JISC

UNIVERSITY OF Hull

History DMP Project

Final Report

Chris Awre, Martin Dow, Richard Green and John Nicholls

April 2012

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.

Table of Contents

Acknowledgements

We gratefully acknowledge funding from the JISC *Managing Research Data* funding programme to support this project.

We are indebted to our colleagues in the Department of History at the University of Hull for their support in this work; clearly, without their willing assistance our task would have been extremely difficult. We acknowledge, too, the work done on our behalf by Martin Dow at Acuity Unlimited in helping us realise a linked data prototype; work, not just in implementation, but in educating the rest of the team about the field. Simon Lamb, in the University's ICT Department undertook the software development new to this project, whilst Jon Higham (also from ICTD) and Andrew Thompson (our programmer on the precursor CLIF Project, now with the SEED unit at the University) took a new look at integration with Sakai,[1] Hull's virtual learning environment. Renzo Sanchez-Silva, a GIS Infrastructure Software Engineer at the University of Virginia, was extremely helpful in assisting us to resolve some questions about GIS coordinate representation in MODS metadata.

Finally, we should acknowledge the help and oversight of David Starkey and Peter Wilson who, with Chris Awre, formed the project's Steering Group.

[1] See: http://sakaiproject.org/

1. Project Summary

The History DMP Project at Hull was the first step in developing a better approach to research data management within the University. Building on the established data management practices of individuals within the Department of History, and on acknowledged 'best practice' guidance, the project worked with researchers to frame a departmental approach. The data management plan (DMP) thus developed will support future research strategy and provide a coherent platform for data sustainability. It is the intention that, post-project, the work done with this particular department will be adapted to other subject areas.

During the course of the project an opportunity arose to work with the team at the Digital Curation Centre's 'DMP Online' service to investigate the possibility of presenting the questions and guidance from the History DMP as an on-line service. The DCC service was investigated but, in the form then available, found not to suit the University's immediate needs. Subsequent testing of a development version of the next generation 'DMP Online' leads us to believe that the new tool will allow us to present our DMPs as on-line templates.

The project also embarked on a number of technical initiatives in order to improve the University's capability to curate, preserve and provide access to research data so as to more fully provide the capability to address technical requirements within specific DMPs. These were to involve providing specialised support for datasets within Hull's digital repository, providing for transfer of datasets to the repository from the institutional virtual learning environment (VLE) – where this is being used as a virtual research environment (VRE) – and exposing one of the repository's datasets as 'linked data' as the first investigation into the requirements and potential benefits of such exposure as part of managing the data. Positive progress was made with all three areas, albeit that implementation of the VRE solution has had to be delayed to beyond the project to fit in with upgrade schedules for the system.

The project concluded that the creation of a DMP is a useful route through which to raise awareness of data management issues. Understanding the requirements has provided a focus for understanding how data management should be best resourced, but also an assurance through the technical work that the institutional repository does have a role to play here.

2. Project Outputs and Outcomes

The History DMP project produced a number of tangible outputs:

- a report detailing the History Department's data management practices prior to the work of the project, the requirements for data management following from these, and a summary of the role of data management in supporting History research. 'The case for a History Data Management Plan' can be found at https://hydra.hull.ac.uk/resources/hull:5121

- working with members of the History Department, and using the Data Curation Centre's *Checklist for a Data Management Plan*[2] (hereafter just DCC Checklist) as a basis, a new DMP for the department was developed. The plan, as at the time of this report, is to be found at Appendix 1; the most up-to-date version can be found at https://hydra.hull.ac.uk/resources/hull:5423
- the History DMP thus created has been tried with three datasets at different stages of development. The results of this work are written up as the 'History DMP Project: Case studies' and can be found at https://hydra.hull.ac.uk/resources/hull:5421
- the University's institutional repository, Hydra,[3] has been enhanced to provide more metadata for and increased functionality around datasets. Provision has been added for geographic and temporal coverage (including visualisation of the geographic coverage), explicit citations, and support for DataCite metadata (including the ability to generate DOIs for datasets)
- an exemplar has been produced to show the potential benefits of exposing datasets as linked data

Less tangible, but no less important have been:

- an increased awareness within the History Department of the need for, and benefits of, better research data management
- the History DMP and accompanying case studies as an exemplar and the basis for improved data management planning elsewhere within and without the University

The original Project Plan for our work envisaged that we would take existing software linking the University's virtual learning environment[4] with the repository from proof-of-concept stage to being production capable. This work would then have allowed the transfer of datasets from the VLE to the repository according to need. In the event, this work will be undertaken post-project; further explanation of this is given at Section 3.3.2.

3. The work of the project

The History DMP project was conceived to build on the established data management practices within the Department of History at the University of Hull. It was intended that it should 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 had been the basis of data management thus far. This departmental DMP would look to support future research strategy and provide a coherent platform for data sustainability in future research. The work to be undertaken would develop an overall plan, and then look to implement this using past, present and future research activities. The role of local technical provision, in the shape of the

[2] The DCC Checklist can be found via http://www.dcc.ac.uk/resources/data-management-plans
[3] See: https://hydra.hull.ac.uk
[4] See: http://sakaiproject.org

University's Fedora-based digital repository, would 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, VRE, and linked data interfaces.

3.1 Informing the work: initial interviews

The first task of the project team was to establish what data management was carried out in the Department of History in the course of current research activities. A series of interviews was arranged with a cross section of staff in which we could explore this issue and in which we could undertake some requirements gathering for the DMP work that we intended to carry out.

In the event some 12 researchers were interviewed and, in addition, the team interviewed several postgraduate students to seek information about their practices and perceived needs.

In choosing the staff to be interviewed the intention was to cover four general areas of interest:

- quantitative data
- qualitative data
- oral testimony
- archival data

Analysis of these interviews allowed us to produce the project's first report *The case for a History Data Management Plan.*[5] Whilst the interviews were undertaken in a relaxed fashion to ensure that the interviewee had the chance to share anything he or she felt might be useful, they were loosely structured around 13 key questions: these form Appendix 4 to this report.

The interviews provided the team with a lot of information about present data management practice. At one end of the spectrum were researchers who had clearly considered their data management needs carefully and who had found quite appropriate ways to address these; at the other end of the spectrum, whilst data management needs might be understood, they had not been implemented in a robust fashion. It will be useful to note here that whilst the project might have been conceived largely with digital data in mind, it rapidly became apparent that interviewees would equally welcome assistance with analogue (paper-based) data as well.

Here is not the place to repeat the contents of our report; suffice it to say that broadly it covered the main areas of the DCC Checklist:

- Data types, formats, standards and capture methods
- Ethics and intellectual property
- Access, data sharing and re-use

[5] Green, Richard A. and Nicholls, John H. (2011) *The case for a History Data Management Plan* University of Hull. At: https://hydra.hull.ac.uk/resources/hull:5121

- Short-term storage and data management
- Deposit and long-term preservation of data

The reader should refer to the report itself for further detail.

3.2 Departmental management plan

The researchers, staff and postgraduate, that we interviewed seemed genuinely pleased that someone was taking an interest in their work and was prepared to help them manage their data in appropriate ways, be those focussed on safe-keeping and/or making it accessible to others. The idea of some sort of departmental (or University) guidance was universally welcomed.

As we suspected prior to the interviews, a number of those we talked to were aware that there were already guides "out there" that could help with an individual's DMP; rather fewer were familiar with their contents. Our History DMP Project took as a starting point for desirable coverage of such a plan the Digital Curation Centre's DCC Checklist (v3.0 17 March 2011). The very few interviewees who knew of it generally considered it too complex for them to find useful. During our interviews it often became clear that individuals were unaware of services existing within the University that might help with their data management and so they were much more enthusiastic about the idea of a 'local' document which more specifically addressed the needs of the department and which could point them at specific services or points of contact; several of them noted that it would be much appreciated if such guidance could deal with paper-based materials as well as digital ones. As an aside, it should not be thought that the DCC checklist was our only source of inspiration; a number of further sources were consulted amongst which Sarah Jones' DCC guide *How to Develop a Data Management and Sharing Plan*,[6] the UK Data Archive's *Managing and Sharing Data*,[7] the DCC's *DMP Online*[8] site and the California Digital Library's online *DMP Tool*.[9]

It became clear in discussions that researchers would welcome guidance that covered management from the first gathering of data through to the end of their research and, further, looked at the issues surrounding the long-term stewardship of the data and whether such material should be accessible to others or not.

3.2.1 Paper version

Whilst colleagues may have felt that the DCC Checklist was too complex for them, it formed a very useful starting point from which the project could develop something that they might find more accessible. We took the approach of using what we considered to be the key 'DCC questions' (given the background of our interviews) and rewording them, where we felt it necessary, in less technical language so that they were more easily understood by our History colleagues. In the current version, individual DCC question themes have often been combined into one and we have

[6] See: http://www.dcc.ac.uk/resources/how-guides/develop-data-plan
[7] At: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
[8] See: https://dmponline.dcc.ac.uk/
[9] See: https://dmp.cdlib.org/

sometimes added additional points of our own. For each non-trivial HDMP question, guidance notes have been provided. Whilst we did not want to create two classes of question, researchers found it useful for us to highlight those questions that would be of particular importance in situations where they might be applying for funding.

The current version of the Hull History DMP can be found at Appendix 1;[10] it is the result of several iterations incorporating suggestions from potential users. The practical process of interviewing researchers and developing several case studies (section 4) resulted in the holistic, organic and evolutionary development of the plan. There is a clear recognition that the plan remains a 'live' document that requires regular revisions to meet evolving requirements; many of these will, though, be factual in nature, for example, the contact information listed in the 'Notes' section. The current plan is envisaged as a template from which later iterations may be developed, and from which further departmental plans (outside the history remit) may be initiated.

3.2.2 On-line version

Whilst working on the paper version of the History DMP we were in contact with colleagues at the DCC in Edinburgh and it was suggested that we might wish to try our plan in their DMP Online tool. The then current version of the tool allowed institution-specific guidance to the DCC questions.

This process involved us in supplying the DCC with a modified version of their (Excel-based) data sheet such that it contained the standard DCC questions but Hull's guidance notes. In order to achieve this we mapped Hull's agglomerations of the DCC questions back onto their original components and simply copied the Hull guidance across. Where the guidance entry is empty the DMP Online tool does not render the question on screen, and so we ended up with the DCC questions deemed most relevant by Hull's researchers accompanied by Hull-specific guidance, albeit presented in 'DCC language' and in more items than we would have wished.

Whilst we were undertaking this investigation we became aware that the DCC were working on a new version of the tool that seemed to offer facilities that matched better the approach we would have preferred. It will allow users to mix disciplinary, institutional and funder templates to put together a set of DMP questions suited to their needs at the time. Unfortunately this new version of DMP Online was still being worked on at the time of this report and so we have been unable to undertake extensive trialling. Such testing as was possible leads us to believe that this next-generation 'DMP Online' will offer the flexibility that we had hoped for and will allow us to construct our own DMP templates for users to complete online, mixing our own questions and guidance with standard DCC provided questions and guidance as required.

[10] This is the version current at the time of this report. The most up-to-date version can be found in the University's institutional repository: https://hydra.hull.ac.uk/resources/hull:5420

3.3 Provision for data online

The HDMP Project was an opportunity not only to consider a data management plan for a University department, but also to review the services available in the University to support such data management in order to better enable implementation; in particular, we had wanted to look at services around data management and curation, learning and teaching, and access to content in linked data form. A number of our interviewees were aware of Hull's institutional, digital repository, Hydra,[11] but were not clear what specific services it could offer them as researchers. In fact, it offers curation of their data in a highly secure environment and can make it accessible to other users or groups in a very granular fashion if and when that is required. Over the longer term, the repository is preservation-capable so as to meet relevant preservation requirements.

3.3.1 Hydra functionality for datasets

The digital repository is based on the Fedora Commons[12] repository software; the primary access to this is via a Hydra[13] user interface (UI). At the time of the HDMP Project Hull's Hydra UI was still under development and lacked specialised provision for dealing with datasets (rather, datasets, a number of which were already held by the repository, were treated as generic files). The HDMP Project gave an opportunity to provide better, more specialised, provision for such content by providing targeted pages on which to create and edit database digital objects and a specialised page format through which to display information about them and from which the databases might be downloaded where security settings allow.

Initial work on such functionality required that we investigate what metadata[14] the repository should hold for databases. The repository uses the MODS schema to record metadata about its digital objects. Many of the metadata fields are standard across the various kinds of repository content but we were concerned to cater for any needs of databases that may go beyond these: whilst the repository could hold more detailed dataset-specific metadata separately (for example, DDI records), we wish to provide variants of the standard MODS-based view onto all content held for ease of use. Four particular issues came up:

- the ability to cope with potentially large numbers of co-authors
- the ability to specify the 'coverage' of a database be that geographical (e.g., the Baltic Sea) or temporal (e.g., 1850-1895)
- where there is geographical coverage to have the option of defining this using values for latitude and longitude and showing it on a map
- the need to be able to describe and cite datasets clearly

[11] See: https://hydra.hull.ac.uk/
[12] See: http://fedora-commons.org/
[13] See: http://projecthydra.org/
[14] Metadata is a word used differently by historians and IT professionals. In this section we are using it in the IT sense.

The full MODS record for a typical Hull dataset object is given as an example at Appendix 2. An example splash (display) page for a Hull dataset object is shown at Appendix 3.

3.3.1.1 Multiple authors

The Hydra repository UI implemented at Hull already provides the ability to add multiple authors to a digital object and sets no limit on the number. The issue that this does raise is that having spent the time once typing all the names for, say, a particular research group it should be possible to save that information as a template. The repository has several other situations requiring the ability to be able to store templates and this adds to the list. It has not been possible to address this issue within the timespan of the HDMP Project, however the learning generated by this particular use case will be addressed post-project to provide a common solution across this and the several other use cases requiring it.

3.3.1.2 Coverage

A number of the existing databases in Hull's repository, including some of those originating from the History Department, have coverage implied in their title. Take, for example, *HMAP Dataset 07: Danish Baltic Catch Data 1611-1920*: it would be useful to expose explicitly in metadata two items of 'coverage' – the geographic coverage 'Baltic Sea' and the temporal coverage '1611-1920'. Software coding for this was a relatively straightforward task and has been completed. Potentially these entries could be used in the repository to provide an additional search facet in the context of databases (or other types of content where coverage is applicable, for instance photographic images).

3.3.1.3 Plotting geographical coverage

When geographic coverage is specified we felt that we could potentially add value by providing functionality which allowed the user to see the coverage (point or region) expressed on a map. In order to achieve this it was necessary for us to represent the appropriate coordinates in the MODS metadata and provide the facility to pass these to the Google Maps tool. Whilst the programming itself was not unduly problematic, it turned out that representing geographical coordinates in MODS was. The current MODS User Guidelines,[15] maintained by the Library of Congress, are short on specific examples of how coordinates might be handled, and older guidelines which have an example seem to portray a methodology best described as perverse. In the event we sought assistance from our Hydra Project partners at the University of Virginia. Their specialist in the area recommended a sensible solution to our problem which combines the basic MODS requirements with the KML guidelines used by Google for representing points and shapes.[16] These we adopted.

Following a number of rounds of prototyping we have provided geographical coverage information at a number of levels of detail ranging from one to three depending on the nature of the dataset.

[15] See: http://www.loc.gov/standards/mods/userguide/
[16] See: https://developers.google.com/kml/documentation/kmlreference#polygon

Figure 1 below shows an example map indicating the approximate coverage of a dataset covering fishing catches in the Southeast Australian shelf.[17]

Figure 1 Example map showing dataset coverage

This map, displayed on the splash page for the dataset is intended to provide no more than an approximate context for the data – this is explained in a paragraph above it (shown in Appendix 3). In the case of this dataset we do have very accurate coordinates for the extent of the fishing area as generally defined. Thus, a link in the splash page download panel provides the facility for a user to download a kml file for use with Google Earth which will render the area in great detail:

Figure 2 Accurate representation of geographical coverage in Google Earth

For those without Google Earth, or not wishing to download the kml file, a link 'View as map' is provided which passes the information to Google Maps in order to render a detailed, but marginally

[17] See: https://hydra.hull.ac.uk/resources/hull:1934

less rich, map in the user's browser (for instance, it can be seen that Tasmania is not delimited as in the Google Earth view – only the outer, bounding, shape is rendered):

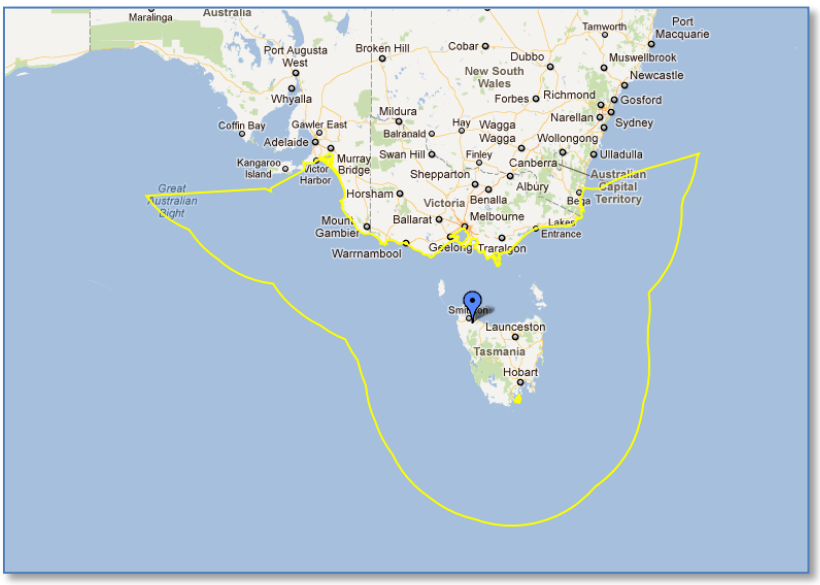


Figure 3 Representation of geographical coverage in user's browser

In all cases the map is fully functional in terms of the normal Google tools for zooming, moving, switching between map and satellite views etc.

3.3.1.4 Dataset description and citation

One organisation attempting international standardisation of database description and citation is DataCite[18], which recommends both a metadata schema and a basic citation structure.

The various options and formats by which scholarly texts can be referenced are well accepted and understood. There is less agreement, particularly across disciplines, about how, precisely, one might go about citing databases. DataCite recommend a citation structure and, importantly, that any citation should contain a linkable, permanent URL in the form of a DOI (digital object identifier).

Whilst attracted to DataCite's suggestions we at Hull needed to accept that there are other ways of doing things and that we needed to provide for them. Thus we will provide for both DataCite and other approaches:

- within our MODS metadata we have made provision for recording a DataCite-style DOI (in a future version we will provide on the MODS editing page for minting a DOI from DataCite for a new dataset through use of the DataCite API – testing for this is ongoing with DataCite at the British Library as part of their development of support for institutional data repositories)

[18] See: http://datacite.org/

- within our MODS metadata we have made available one (or more) fields for recording entire citations in any manner required by the dataset depositor
- we have provided an XSLT transform that converts our MODS metadata to DataCite format metadata 'on-the-fly' in response to a user request. (We already successfully use this technique for a number of other metadata transforms.) Such DataCite metadata can be generated by clicking a link on the dataset's display page; the link is only present when a DOI exists (mandatory in DataCite metadata).

Thus provided for we hope that we can be very flexible in our future dealings with datasets.

3.3.2 Sakai/Hydra link

A previous project undertaken by Hull and King's College London was CLIF,[19] the Content Lifecycle Integration Framework. Funded under the JISC *Repositories Enhancement* strand, amongst other things this work sought to enable two-way transfer of files between the University's virtual learning environment, Sakai, and the institutional repository. During this work software was developed to a 'proof-of-concept' level to achieve this functionality. It had been the intention of this project to take that work and raise it to production status so that datasets in the repository could easily be transferred into Sakai to aid learning and teaching whilst datasets developed in the Sakai environment could be transferred to the repository for curation and preservation.

Discussions to inform this part of the project highlighted three aspects that required attention:

- The proof-of-concept demonstrator was not integrated fully with the security mechanisms in place on the repository and achieving this integration would be a significant piece of work in its own right to enable full equivalence of data management across the two systems.
- This focused discussion on exactly what we wished to achieve through this integration. The use case that emerged limited the transfer to those datasets researchers were generating during their research, and wishing to archive into the repository (i.e., to capture works in progress). This would then be a focused transfer into a specific part of the repository, on which the security could be more easily managed.
- As plans progressed, parallel VLE management proposed a major upgrade to the Sakai software (to version 2.9). Because of the nature of the integration, it seemed likely that any work we carried out on the current version would have to be repeated once the new version of the VLE was in place. Clearly, this would not have been sensible use of time (although some preliminary investigation has taken place).

Taking these aspects into consideration, it was decided to focus on the specific use case identified, but delay the work until the upgrade path had been clarified so as to maximise the benefit of undertaking it. This will be post-project, most likely during the summer vacation period.

[19] The CLIF Final Report can be found at https://hydra.hull.ac.uk/resources/hull:4194

3.3.3 Linked data

The final area of the technical work undertaken by the project was to consider the place of linked data in the University's provision. This is an area in which the University-based members of the team had little direct experience and were thus fortunate to be able to work with Acuity Unlimited in order to explore it.

Linked data is about applying the principles of the web to sharing data, and doing so at a deeper level than just making a monolithic data file available on the web. Instead, each 'thing', or entity, within the data is given an individual identity and ascribed an identifier, a URI.[20] This confers several properties to the data:

- *the data is placed in context:* each item of data has a web address through which it can be annotated and referenced, this allows explanations and implications to be linked back directly to the data, at a fine-grained level;
- *the data is linked to its information model*: well-structured data adheres to an information model or schema to help describe it – an "ontology" with vocabulary terms. Each element of schema information used in modelling the data is also given a URI allowing data queries to be formulated and automatic inferences made based on the schema to form links across datasets; information can be combined across silos, enhanced by combining it with third party data sources and further contextualised;
- *the data is accessible 'live' at a fine grain* – as this data is accessible over the web in a decentralised fashion, downstream applications have the flexibility to run from the 'live' data on the web as an alternative or in addition to using static data dumps cached locally.

Put another way, the purpose of exposing linked data is to share information in such a way that remote queries can be undertaken not just within a single dataset but across multiple datasets from different providers quite possibly located in different places around the world. Remote data may be mixed with local data using the same tools and methods. Managing data in this way allows many possibilities for its use: visualisation, analysis, collection, sharing, enrichment, data mining, publication, and much more…

Clearly, for this approach to work it is necessary that there be standard ways of presenting and describing the data. The web of data is created using HTTP,[21] URIs and RDF[22] and the following four 'design principles' articulated by Tim Berners-Lee:[23]

1. *Use URIs as names for things*
2. *Use HTTP URIs so that people can look up those names*

[20] A URI is a string of characters used to identify a resource allowing interaction with it (here, over the web, by means of HTTP URIs)
[21] HTTP: The Hypertext Transfer Protocol is the basis for data communication on the world wide web
[22] RDF is the World Wide Web Consortium's 'Resource Description Framework'
[23] Berners-Lee, Tim (2006) *Linked Data* at http://www.w3.org/DesignIssues/LinkedData

3. *When someone looks up a URI, provide useful information, using the standards (primarily RDF, SPARQL[24])*
4. *Include links to other URIs, so that they can discover more things*

Publishing information using the above design principles allows others to reuse it and create innovative applications. The power of the approach is derived through the way the web works; it is highly decentralised which leads to highly decoupled data flows. No central body determines the "correctness" of combined data or orchestrates the integration processes. Within this environment if there is not to be the potential for ambiguity, it is, however, necessary to have agreed ways of describing your data (using RDF structures and an ontology language[25]) and agreement about the use of terms (standardised vocabularies) and their conceptual relationship with each other (the ontologies themselves). It is necessary that, having data referencing the 'North Sea', one can know with confidence that one is talking about the large body of salt water rather than an ocean-going vessel of the same name. Equally one needs to express that the body of water that forms the 'North Sea' is one and the same entity (real world thing) as referenced by the label 'La mer du nord' in a French database.

In order better to understand the potential for linked data in the context of History data management the project took an existing, published dataset from Hull's Maritime Historical Studies Centre and exposed it on the web as linked data. A simple demonstrator was put together to show how this data could usefully be queried and, perhaps more importantly, how the use of linked data allowed discovery of useful, related information beyond the original dataset. The steps necessary to achieve this can be summarised in the following manner:

1. *Data awareness* – what is the data we have and what standardised vocabularies might we use to describe it?

2. *Data modelling* – what tools and processes are there available that will assist in the construction of data structures, vocabularies and conceptual ontologies?

- Data cataloguing - modelling the metadata about the dataset for use in discovery processes
- Data type information and ontology linking - aligning terms and ontologies from differing sources
- Statistical and numerical data - how can numerical data be modelled for downstream reuse?

[24] SPARQL is a query language widely used against relationships expressed in RDF

[25] The most common ontology language is the Web Ontology Language, OWL - see http://www.w3.org/TR/owl2-overview/; also see RDF-Schema, RDF-S http://www.w3.org/TR/rdf-schema/ and Semantic Web Rule Language, SWRL http://www.w3.org/Submission/SWRL/. Thesaurus structures and authority files expressed in SKOS may be considered conceptual schema that contain entities - See http://www.w3.org/TR/skos-reference; for relevant example see VIAFontology - http://viaf.org/ontology/1.1/

3. *Conversion for publishing* – the data needs converting to usable RDF

- sharing datasets by means of using machine-readable catalogues
- data cleanup, where necessary
- "reconciliation" to find URIs for entities found within the data and consistent allocation of URIs for the same entity
- conversion to RDF structures and interlinking between these
- Application of appropriate licence terms
- Allocation of persistent identifier(s) for dataset - refers to a step the institution may wish to take that provides guarantees over the stewardship of data at a URI, such as registering a DataCite DOI[26] or referencing a 'PID' in the Hydra repository
- Potential steps towards including linked data artefacts within repository lifecycles

4. *Discovery* – how do we render our data discoverable based on differing methods and criteria?

- dataset metadata – data catalogue vocabularies VoID[27] and DCAT[28] assist discovery based on information about the contents of the dataset
- semantic search – for example, Sindice[29] is a web crawler and search index that provides discovery service for RDF data, so certain minimal indexing requirements should be met. Mainstream search engines also offer support.[30]
- CKAN[31] is an example of a global registry that provides a destination and discovery point for datasets
- Local hosting options – implementing content negotiation to resolve URIs equally for humans looking for web pages and machines requesting RDF,[32],[33] and locally maintaining a SPARQL end-point to support queries of locally-hosted RDF data
- Dedicated cloud hosting options – e.g., Kasabi[34] - provide services assisting data discovery

[26] See: http://datacite.org/
[27] A W3C Note: see http://www.w3.org/TR/void
[28] Focus on cataloguing government data: status at http://www.w3.org/2011/gld/wiki/Data_Catalog_Vocabulary
[29] See: http://sindice.com
[30] schema.org vocabulary, see http://www.schema.org
[31] Portal branded the Data Hub: http://thedatahub.org/
[32] A linked data practice to resolve HTTP URIs to metadata about the URI referent. For suggested implementations from the W3C see: http://www.w3.org/TR/cooluris/#implementation
[33] The case study comments on an important difference between key assumptions that are made by linked data (via the W3C Technical Architecture) and repository architectures as to the significance of the "representation" of a "resource" and what is therefore returned by a given URI.
[34] Talis have established Kasabi as an information marketplace in the cloud based on linked data - see http://kasabi.com/

5. *Integration and aggregation* – how do we bring together our data and related data from elsewhere? Support varies in nature, for example:

- LATC European data integration service funds support for 'consolidation' tools[35]
- Sig.ma[36] – finds data within datasets based on multiple index entry points (OKKAM[37] entity base and SPARQL query endpoints)

6. *Application* – what do we want to be able to demonstrate to derive value from our linked data? Various elements come into play:

- "Meshups not mashups" – local data is not privileged, generic applications can be adapted to consume various "meshups" of data rather than "mashups" of software application parts.
- Software application features should be tailored to be specific to the use case required of the data
- Construction of user-defined data subsets as a value-adding device
- Faceted search interfaces could potentially be lightweight to construct
- Enrichment of local data sources with data sourced from the web, e.g., for background information or including information specific to a related discipline
- Some UIs are available as "templates" to help simplify development[38]

The actual process of carrying out this work, and the detailed research and thinking involved, is too complex to include here but will be made available as a document in its own right in early summer 2012.[39]

4. Case studies

The project used the Data Management Plan described in section 3.2 as the basis for three case studies.[40] These applied the Plan to three dataset scenarios:

- an existing dataset
- a dataset associated with ongoing research
- a dataset created in conjunction with a new research project

In each case, researchers were asked to complete the questionnaire and were then asked about

[35] For examples of vocabulary and identity consolidation, see http://www4.wiwiss.fu-berlin.de/latc/toollibrary/categories.php#consolidation
[36] See: http://sig.ma for web service and software download
[37] See: http://okkam.org
[38] For example, see the Simile Exhibit project, http://www.simile-widgets.org/exhibit/
[39] The document will be made available in Hull's institutional repository at: https://hydra.hull.ac.uk/resources/hull:5461 in due course
[40] The case studies and covering documentation can be downloaded from https://hydra.hull.ac.uk/resources/hull:5421

- its ease of use
- its usefulness as a checklist against their normal data management routines
- its usefulness in making them think about data management issues they might not otherwise have considered
- the usefulness of the accompanying notes in identifying services available to them to assist data management and that they might not otherwise have known about or considered

Feedback from the researchers was used to modify and improve the Plan and notes.

The case studies document outlines the method, findings and conclusions derived from the process of actually engaging several researchers for the purposes of incorporating 'real life' scenarios into the HDMP.

The primary conclusions drawn were:

- The HDMP is relevant and useful across a full range of research ranging from individual to interdepartmental, national and multinational projects, and spanning a multitude of historical and related research approaches and subjects.

- The importance of providing a Data Management Plan is recognised as an essential element of any successful project, particularly in terms of establishing a coherent strategy for the successful completion of a project.

- Our interviews revealed a vital aspect of the entire process: historians are not necessarily willing or able to address issues that are outside the scope of their personal remit (i.e. doing history). However, it is clear that the importance of addressing these issues was recognised, and it was accepted that the levels of assistance (via the accompanying notes and contact details) did provide a clear and adequate means of addressing the issues.

- Of particular interest was the idea that the plan is a time saving and useful tool in the overall business of historical research procedures.

- Each of the case studies identified that the History Data Management Plan is a simple yet effective tool that assists in the strategic planning process and highlights aspects of work that may otherwise be overlooked, or left until they become very difficult to manage. For example, careful consideration and planning for the future storage and availability of data after the project lifespan has expired must be undertaken in good time.

- The case studies reinforced our view that the History Data Management Plan is a perpetual work in progress and that it will change in time. Details, such as contact information, will need to be updated regularly to ensure that the plan remains a useful and workable tool in the future.

5. Project achievements

The time working on this HDMP Project has provided a useful learning exercise for all involved. Starting from the situation where Hull's History Department had myriad approaches to data management, we have arrived in six months at a potential unified approach effectively developed by and thus supported by its researchers themselves. Clearly this alone benefits the members of the History Department but in addition there have been wider reaching benefits. Researchers in the department have become aware, or have been reminded, of services and individuals within and without the University that may be a source of support for data management. In addition, those responsible for providing some of the University's support services have had the chance to refine and extend their provision for assisting the department. The members of the HDMP team from Library and Learning Innovation are confident that the approach taken, and lessons learned, here will be broadly applicable when working with other departments.

Key to the success of this project, and a key learning outcome, has been the involvement of the researchers themselves. In gathering their requirements at the start we not only became informed on the issues that needed addressing through a data management plan, but also informed those being interviewed that managing data was not something they needed to do by themselves, particularly when it was an area that lay outside their comfort zone as academics. In developing case studies in the use of the DMP, the project demonstrated that what had been produced did work, and will be a valuable asset in supporting future data management, both in History and other departments.

We had initially started out thinking we would adapt the DCC checklist for departmental use in our DMP, mainly because it was felt that the detailed checklist needed to be presented in a simpler fashion to help engage with the academics and not put them off. We have indeed simplified it, and this has proved successful. However, in highlighting this approach in an early deliverable on user requirements, we engaged with the DCC on how we could translate our DMP for use with the expanded checklist in the context of the DMP Online tool. The project has run in parallel with developments in this tool, but the outcome appears to present us with an option when providing the DMP: document or online.

Technically, the work to develop Hydra to better provide for datasets has proceeded smoothly. We anticipated that it would not be complex on the basis that Hydra has been designed to be adapted for such circumstances (we shall be working on images later in the year), but it was nevertheless pleasing that the work did turn out as it did. We will be continuing our investigations on the use of DataCite beyond the project as the BL's service engages further with institutional repositories. The linked data work undertaken has provided additional stepping-stones towards understanding how we might exploit this approach in managing datasets for better exploitation. The link from the local VRE to the repository has been affected by upgrade timetables, though the clarity around the use case developed will now inform this work more precisely.

6. Immediate Impact

6.1 Immediate impact on History department:

- Raised awareness of data management as a relevant and necessary aspect of research practice – researchers specifically commented on their expanded understanding of the procedures that they need to follow in order to see the successful development of their research projects; this has been further evidenced by the increased level of requests for assistance regarding data management aspects of projects, particularly in terms of data storage and availability.
- Management willingness to participate – head of department, head of research and relevant staff members are incorporating the DMP and the fundamental principles that underpin it into departmental practice; new projects, ranging from very large (multi-institutional/departmental) to very small (individual and/or PhD thesis) will all employ the DMP in the initial phase of development.
- The History department was invited to contribute to a bid for ongoing development of data management materials in conjunction with the Institute for Historical Research and the UK Data Archive. The work of the current project is benefitting from dissemination through these national bodies to inform the history community.

6.2 Immediate impact on the University:

- The creation of the DMP and the learning from carrying out this work has enabled Library and Learning Innovation to take the lead in developing the roadmap for meeting the EPSRC's expectations around research data management. This roadmap has placed data management planning in the wider context of what is required to support research data management and dissemination more widely.
- The work of the project will inform an institutional engagement with the Digital Curation Centre to raise the profile of data management within the University. This is scheduled to take place during the summer of 2012.
- The local mechanisms through which data can be managed locally are now more firmly established, primarily related to the repository. This has facilitated conversations with other departments on how they might develop their data management.

7. Future Impact

- It is anticipated that the work of the project will have an impact on a range of other departments across the University through adaptation of the existing plan. The Faculty of Health and Social Care has already expressed interest and further dissemination is planned. The departments/faculties adopting it will be tracked by Library and Learning Innovation as part of the dissemination activity.
- Dissemination to the academic history community will be continued, with the intention of both raising awareness of data management within history and also increasing usage of the DMP. The DMP itself is being made available through our repository, and we can track the

number of downloads for the file. Whilst not an exact measure of take-up, it will enable us to gauge interest and allow appropriate follow-up.

- The role of Library and Learning Innovation within this project has placed it at the hub of research data management for the University, especially as there is no dedicated research support staff centrally. To what extent this is developed as a role for the Library remains to be seen, but we are now well placed to address a range of data management issues.

8. Conclusions

The majority of the conclusions that emerged from the case studies, as listed in section 4, can also be applied to the project as a whole. General conclusions from the project in addition to this are:

- The DMP approach to understanding data management has been a valuable one. We had established records for a number of datasets in the repository before undertaking this work, but admittedly on an ad hoc basis as we were requested to do so. The DMP approach, particularly at a departmental level, has provided insights into how we should be managing data, and how we can support academics in addressing the various issues that arise when working with data.
- The resourcing of local data management capability will be a continuing challenge as DMPs are taken up and demands increase. Involvement from within the History department that was technically aware was invaluable in making this project a success: not all departments will be in the same position. Likewise, we have been fortunate to have technical input to those areas we have addressed that will not always be readily available. Nevertheless, developing the DMP has provided us with the wherewithal to make the case for resources as required to enable the advantages that can come from good data management.
- Our institutional repository can be used as a data repository alongside more textual collections. Whilst the Fedora/Hydra infrastructure was designed with this in mind, putting it into practice has demonstrated how the design has been made real.
- History has been a valuable subject within which to work. It does not have the complexity or the amount of data some science disciplines do, but it does have a breadth of activity that encompasses a wide range of data issues, and so provides a useful discipline within which to explore data management issues.

9. Recommendations

The following recommendations are made on the basis of the project's work:

- That establishing a data management plan at an early start of the research process will pay dividends later on. Nevertheless, research projects at all stages would benefit from adopting a DMP to guide their work with data.
- That institutions look to their existing institutional repositories to understand to what extent they can be used to support data management and should not assume it requires a different system. Institutional repositories may need work, but locating datasets alongside other

research materials will facilitate access and the sharing that is encouraged. It is recognised, of course, that not all such repositories will be in a position to do this, and that all datasets may not suit such an environment. However, there is value in considering use of the institutional repository as an option.

- That champions are sought within different disciplines to provide advocacy within a department/faculty and facilitate engagement with academics.
- That DMPs are regularly reviewed to ensure they do not become stale or out of date as data management requirements evolve and local or other support changes.
- That institutions recognise what they can provide locally and what services they need to source externally. This suggests the possible benefit of a DMP development plan in terms of the supporting infrastructure that can be provided.

10. Implications for the future

A number of implications for the future are encapsulated within the previous sections and are not repeated here. However, additionally we anticipate the following:

- Noting the work in other projects on institutional data management policy, we need to take account of these in developing our data management capability starting from History. It will be of interest to contrast the impact of an institutional approach against a discipline approach as a starting point, and it will be important to identify where they meet in the middle.
- We shall be feeding back our experiences to the Hydra community to facilitate other Hydra adopters in their work on data management. The software used within the project is available through github[41] and we welcome further collaboration on additional work. We have found Hydra to be a flexible interface solution to make the most of the Fedora functionality underneath, and recommend its investigation by others.
- The role of Library and Learning Innovation, and the repository activity within this, may be altered by this growing involvement in data management. This requires a shift in perspective and skill sets that will need to be addressed to be able to provide the best service we can.

[41] See: https://github.com/uohull/hyhull

Appendix 1: Hull History DMP

Note that page numbers in the 'Table of Contents' relate to pages in the stand-alone document, not to pages in this report. The area available for responses has sometimes been reduced here in order to save space.

This is the version of the DMP that was current at the time of this report; the most up-to-date version can be found at https://hydra.hull.ac.uk/resources/hull:5423

History Data Management Plan
Department of History
University of Hull

Current Date	
Researcher Name(s)	
Project Title	
Project Description	

Important Notes:

1. For detailed, updated explanations of the various parts of the document that require completion, please refer to the accompanying Appendix.
2. **ALL** questions should be answered; shaded areas of the document indicate areas that specifically apply to the types of questions relevant to funding criteria and may prove very useful when submitting bids or making applications.
3. This University of Hull History Data Management Plan (HDMP) applies the DCC Checklist for Data Management (v3.0 17 March 2011).

Contents

Section 1: Project Information

1.1	Project Name / Title	1.1.1
1.2	Project Duration (From–To)	1.1.4
1.3	Partners (where applicable) – list primary first	1.1.5 1.1.6
1.4	Brief Description (Aims and Purpose)	1.2
1.5	Historical Period researched (where applicable)	HUL 1.2
1.6	Departmental/University requirements (where applicable)	HUL 1.3.2 1.3.3
1.7	Funding Body or Bodies (where applicable)	1.1.2
1.8	Budget (where applicable; estimate if not established)	1.1.3
1.9	Funding Body Requirements for a Data Management Plan (where applicable)	1.3.1

Section 2: Data, Materials, Resource Collection Information

2.1	Brief Description of the Historical Data, Materials, Resources being researched, collected and collated: (e.g. Shipping Log Records 1890-1900; Voice Recordings factory Workers 1950-1980; etc.)	2.1 HUL
2.2	Strategy for Collation of Data and Materials (e.g. Access to archive requested, Library materials borrowed, photocopies made of documents, volunteers enlisted, etc.)	2.2.3 HUL
2.3	Is this unique data? *(NOTE: Both options may be applicable in certain circumstances; analysis and informed composition of existing data may be regarded as "unique")* *If **YES** - provide a brief description of the process for gathering the data (e.g. recorded interviews, compiled statistics, transcribed documents, etc.)* *If **NO** – provide a brief description of the existing data (e.g. documents, database, recordings, etc.)*	2.2 2.3 HUL

2.4	Are there any existing forms of data that may be added to, attached as appendices, incorporated into the data, etc.? If yes, provide a brief description (e.g. 1820-1840 compiled in *A.N.Other* – 1841-1850 will be added).	2.2.1 2.2.2
2.5	Will the data be digitised or appear in digital form (e.g. dataset/database, spreadsheet, PDF Documents, WAV sound files, JPG image file, etc.)? If yes, specify the format(s) that apply.	2.3.2 2.3.3 HUL
2.6	Will the data be non-digital or appear in physical form (e.g. paper document, physical model, printed photograph, microfilm, archaeological artefacts, etc.)? If yes, specify the format(s) that apply.	HUL 2.3.3
2.7	Will the data be stand-alone, or will it be accompanied by explanatory documentation (e.g. metadata, descriptive documents, referenced printed work, etc.)?	2.5.1 2.5.2 2.5.3
2.8	What Added Value will the data bring to the History Community?	2.4.3
2.9	What steps will be taken to assure Quality Assurance/Management (e.g. Monthly supervised checks on data quality and consistency, etc.)	HUL 2.5.5

Section 3: Ethics, Intellectual Property, Citation

3.1	Are there any Ethical and/or Privacy issues that may arise from sharing your data, or parts of it? If yes, provide a brief explanation.	3.1.1 3.1.2
3.2	Will the data comply with applicable British and International Law and what will be done to ensure compliance (e.g. Private information protected under the Data Protection Act, Copyright Legislation, Intellectual Property, checked with supervisor/legal department, etc.)?	3.1.3 3.1.4 3.2 HUL
3.3	If the project is a multi-partner or multi-participant research process, what will be produced to ensure that applicable Law and regulation are adhered to (e.g. joint working methodology, mediation process, etc.)	3.2.4 HUL

Section 4: Access and Use of Information

4.1	Are you required to, or do you intend to, share the information you have researched at any stage? If yes, stipulate when (i.e. upon completion of project, at a specific stage, etc.)	4.1 HUL
4.2	If your information is to be shared, how will this be achieved (e.g. online, printed copy, hard copy, etc.)?	4.1.3 HUL
4.3	Will the information need to be stored for specific/limited use only at any stage (e.g. funder requires access at mid-point review and at end of project when it will be released, etc.)?	HUL 4.1
4.4	Who will be interested in or will require access to your data (e.g. students, funding body, public, history community, etc.)? Stipulate more than one option if necessary.	4.3.1
4.5	How do you anticipate your information being used once made available (e.g. students may use for specific module, history community may use for informing specific research area, etc.)? Stipulate more than one option if necessary.	4.3.2

Section 5: Storage and Backup of Data

5.1	Where will the information be stored **during the lifespan** of the project (e.g. University Server, DVD media, local hard drive, filing cabinet, etc.)?	5.1.1
5.2	Where will the information be stored **upon completion and after completion** of the project (e.g. University Server, Online provision, accessible hard copy, etc.)?	HUL 5.1
5.3	What provision is being made for regular backups or making safe of the information being researched (e.g. extra copies made, a variety of media/methodologies employed, CD/DVD, Data Drive, printed documents, etc.)?	5.1.2 5.1.3 5.2.1 5.2.2 HUL
5.4	Will different versions of the information be stored (e.g. weekly updates will include latest version clearly named as current version, etc.)?	5.3.3 HUL

Section 6: Archiving and Future Proofing of Information

6.1	What is the long-term strategy for preserving the information resulting from the research (e.g. database to be made publically available via University's online provision)?	6.1 HUL
6.2	Will the information be kept after the life of the project? If YES, then for how long, and in what format (e.g. database will be made publically available for at least 5 years after project completion, etc.)?	6.2 HUL
6.3	If the researched information includes sensitive or confidential data, how will this data be managed (e.g. specific data will not be made publically available, sensitive information will be deleted upon completion, etc.)?	HUL
6.4	If the researched information (i.e. dataset, database, documents, etc.) will be accompanied by explanatory documentation or *metadata*, how will these be linked and preserved alongside the researched information (e.g. stored in same process as researched information, stored separately, etc.)?	6.3 HUL
6.5	How will issues of citing and/or referencing the research information be addressed (e.g. specific citation information included, no citation information available, etc.)	HUL 6.3.5

Section 7: Resourcing of Data Management

7.1	List the specific staff tasked with the role of implementing and carrying out the data management aspect of this project (e.g. Dr A.N. Other, Research Fellow, Department of History, University of Hull).	7.1 HUL
7.2	What is the funding strategy for Data Management **during** the project lifespan (e.g. backup materials to be purchased at researcher's expense, project specific ICT equipment to be purchased [give details], etc.)?	7.2 HUL
7.3	What is the funding strategy for Data Management **after** the project lifespan (e.g. University to continue archiving process, equipment to be regularly updated for 5 years after project ends, etc.)?	7.3 HUL

Section 8: Review of Data Management process

8.1	How will this Data Management Plan be adhered to (i.e. what means are provided, such as checklists, regular revisiting of the various elements on the plan, etc. to demonstrate that the plan is being regularly reviewed)?	8.1 HUL
8.2	Who will carry out the regular reviews of the Data Management Plan (e.g. Dr A.N. Other, Research Fellow, Department of History, University of Hull)?	8.1 HUL

Section 9: Statements and Personnel Details

9.1 Statement of Agreement

I/we agree to the specific elements of the plan as outlined:

Project Lead or Individual Researcher:

Title	
Designation	
Name	
Date	
Signature	

Project Supervisor or PhD Supervisor:

Title	
Designation	
Name	
Date	
Signature	

Project Researcher(s) *

Title	
Designation	
Name	
Date	
Signature	

* More than one Researcher may be involved. Continue on a separate page if necessary.

9.2 Expertise of Researchers ***

Title	
Name	
Contact Details	
Expertise	

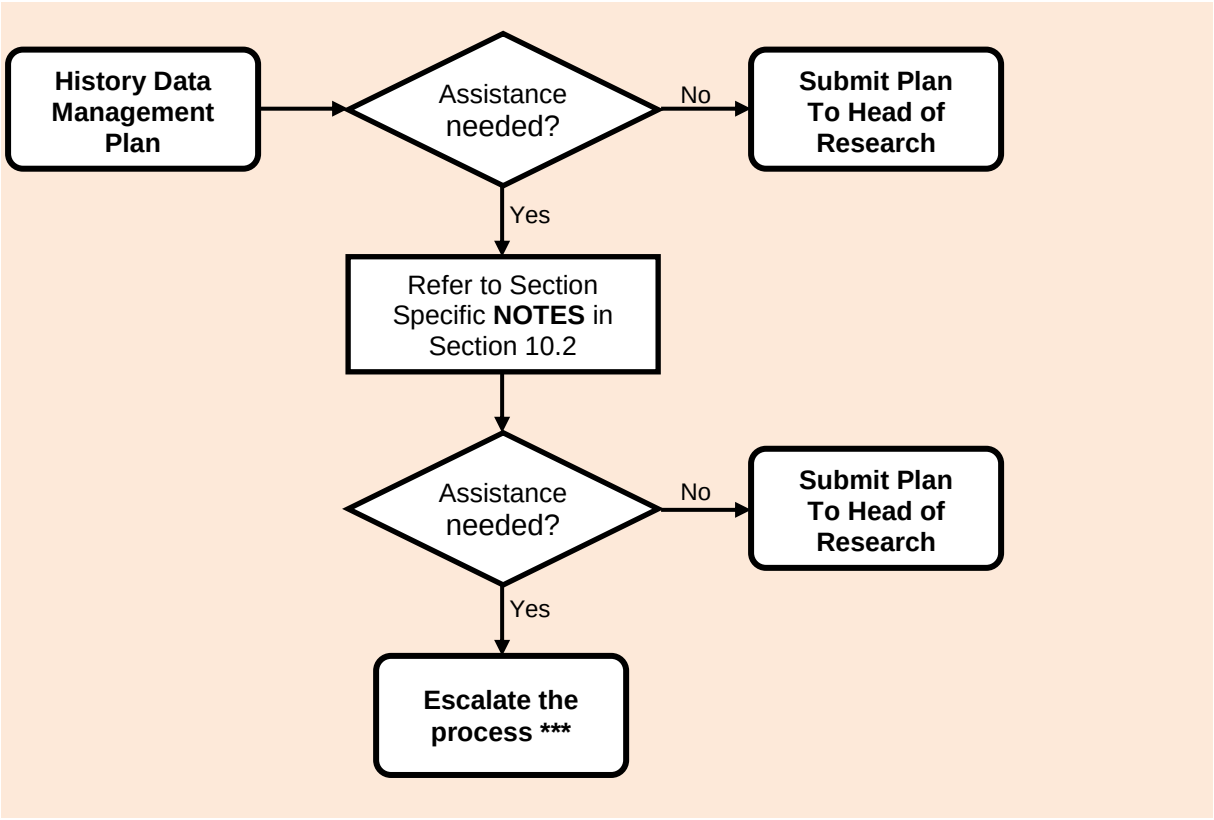
Title	
Name	
Contact Details	
Expertise	

** More than one Researcher may be involved. Continue on a separate page if necessary.

Section 10: Appendices

10.1 Specific Help with completing the Plan

In certain instances, specific guidance may be required in order to complete this Data Management Plan. Assistance should be sought by following the flow chart below:



Escalate the process by requesting assistance from the Departmental Head of Research. Typically this will entail contacting the Data Manager, IT Services and/or Library Services. Specific assistance may be available through the Research Office as well.

10.2 Notes

These notes refer to the specified sections and subsections in this document. Any areas not addressed may be referred to the project lead, supervisor, or the Head of Research. Technical issues may be addressed to the HDMP development team in the first instance.

Front Cover

Details are required to ensure the correct future referencing, storage and archiving of the Data Management Plan. There will be strict adherence to applicable law, including the Data Protection Act; this information will not be made available outside of the specific remit of the History Department of the University of Hull.

Section 1: Project Information

1.1 No specific guidance available
1.2 No specific guidance available
1.3 Required for funded projects – in the event of a single applicant, the primary is the applicant.
1.4 If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text.
1.5 No specific guidance available
1.6 Details may be requested from the project Supervisor, or the Head of Research.
1.7 No specific guidance available
1.8 Applies specifically to funded projects. If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text.
1.9 Applies specifically to funded projects. If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text. Details may be requested from the project Supervisor, or the Head of Research.

Section 2: Data, Materials, Resource Collection Information

2.1 If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text. NOTE: details may change as the project evolves; provide a best estimate.
2.2 If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text.
2.3 If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text.
2.4 “Existing data” may be in the form of an already published data resource (for example, an online database, library items, artefacts, etc.). If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text.

2.5 It is vital that there is a clear understanding of exactly which data types are being discussed in order to plan for future storage, accessibility and integrity. Example data types and formats are available at http://en.wikipedia.org/wiki/List_of_file_formats.

2.6 A great deal of non-digital data may need to be stored securely and/or archived. Various examples of this type of data are:

- Documents: Printed digital, Original artefact, , etc.
- Images: Photographs (size, print type, age), posters, etc.
- Artefacts: Physical model (scale/non-scale, size, availability), archaeological, etc.
- Film: 8/16/32mm, Video, microfilm, negative, etc.
- Other: Live performance, logical model, etc.

2.7 "Standalone" implies a provided information resource that requires no further explanation and may be utilised "as is" without additional resource. Accompanied implies information that is informed by accompanying documentation or resource(s) which help to understand the resource. For example, a database may need to be accompanied by a "metadata" informative document which explains the purpose, use of specific fields, and instructions for utilisation. Details may be requested from the project Supervisor, or the Head of Research.

2.8 This item draws on the fundamental area of historical enquiry; what is the historiographical impetus, or historical purpose behind the provision of this information/research, and how may it benefit the historical community. Generally this will either be the full or a paraphrased version of the abstract for a piece of research. Details may be requested from the project Supervisor, or the Head of Research.

2.9 Quality Assurance/Management in this context refers to the concise provision of a breakdown of what will be done to ensure that the project's progress will be monitored for accuracy, quality of work or research, and timely delivery at regular intervals. Typically, this would be the remit of the Research Supervisor, the Project Lead, or the Head of Department. Details may be requested from the project Supervisor, or the Head of Research.

Section 3: Ethics, Intellectual Property, Citation

3.1 If your research impacts on the welfare, confidentiality or economic status of any individual or corporate group, this should be clearly stated. This scenario is unlikely in historical research, but may arise in instances where individuals reveal personal, confidential or sensitive data; also, it may not be appropriate to reveal certain details in the public domain in the interests of individuals or corporate groups. If necessary, further information may be provided on an attached, clearly labelled **typed** or **printed** sheet. For online forms, the space will automatically be increased to accommodate extra text. **NOTE**: details may change as the project evolves; provide a best estimate.

3.2 It is vital to comply with applicable law. Provide a brief outline of how relevant legislation and regulations will be complied with where appropriate. Where there is any doubt, the first line of contact is the project Supervisor, or the Head of Research.

3.3 See note 3.2 above. Partners in the project must be held to the same legal and regulatory standards. Partners are also protected by applicable law and may avail themselves of the prospect of legal recourse in the event of any perceived illegality or infringement by any party. This applies to all participants effecting or affected by the research project. Where there is any doubt, the first line of contact is the project Supervisor, or the Head of Research.

Section 4: Access and Use of Information

4.1 Sharing data, i.e. making it publically available, may be a requirement of a funding bid, or of a University research project (e.g. Doctoral thesis or research project). Details may be requested from the project Supervisor, or the Head of Research.

4.2 Provide details of how you intend to share your data (if relevant). This may include several options, such as an online accessible dataset or database, or online images. It could also be in the form of a paper based document or set of documents. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

4.3 If your data is sensitive (e.g. not suitable for general access until you have completed, or contains personal data or information) you may need to keep the data secure until you are ready to publish – if at all. Similarly, if the project funder requires "mile-stone" releases, this should be flagged. If in doubt, check this with the project Supervisor, or the Head of Research.

4.4 It is vital that you have a clear perspective of who the outcome of your research is intended to reach. Funding bodies may stipulate specific outcomes – e.g. public access, etc.

4.5 Funding bodies will typically require an explanation of the usefulness of your research once completed, and you should be able to provide a clear idea of what will be done with your data once published or released. Certain obvious options should not be overlooked, such as: paper presented at conference for history community, or book chapter published for community and public research/interest, etc.

Section 5: Storage and Backup of Data

5.1 It is vital that the research materials and data are kept *safely at every stage* of the research process lifespan. There may be help available from IT Services, the Library or the Department. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

5.2 As for 5.1 above, it is vital that you have a clear understanding of how, where and when the research materials and data will be maintained after research process lifespan. This is particularly true where funding bodies have specific outcome criteria (e.g. making a public website available, etc.). There may be help available from IT Services, the Library or the Department. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

5.3 Similarly to 5.1 and 5.2 above, it is vital that you have a clear understanding of how, where and when the research materials and data will be backed up and kept safely, both during and after the after the research process lifespan. This is particularly true where funding bodies have specific outcome criteria (e.g. ensuring that online datasets are maintained for a specific period after the end of a project, etc.). There may be help available from IT Services, the Library or the Department. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

5.4 Very often work is added to, revised or altered and older versions are either overwritten, left as they were, or deleted. It may be wise to maintain a clearly labelled and stored set of older versions of current work in order to backtrack if necessary. It is imperative that a logical and sequenced filing system is used. On computer systems this may be attained by uniquely numbering each version. A useful means of achieving this is by using the current date and time as the unique numbering reference – e.g. "yyyymmdd History Data Management Plan".

Section 6: Archiving and Future Proofing of Information

6.1 Provide information about how you intend for the project outcome(s) or deliverable(s) to be maintained after the end of the project. For example, a dataset may be perpetually maintained by the University's online provision. However, this will need to be confirmed. There may be help available from IT Services, the Library or the Department. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

6.2 Any information that is kept after the lifespan of a project will still need to be stored safely, maintained and be provided in a useable format. If specific file formats are used, they may become unusable after a few years as new software replaces the old. Also, media such as DVDs, CDs and diskettes may become unusable after a while. There may be help available from IT Services, the Library or the Department. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

6.3 It is vital that any confidential data (e.g. personal information about any individual who is protected under the terms of the Data Protection Act, or information that may infringe copyright if released, etc.) must be kept and maintained in a secure environment. All reasonable steps should be taken to ensure the safety of such information. This applies to any information that is kept after the lifespan of a project as well. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

6.4 Datasets, databases, standalone documents, and even artefacts may prove useless without explanatory notes (metadata) accompanying them. These materials need to be clearly linked to the materials so that they can adequately inform any future user about the material. For example, a published dataset will typically be accompanied by a metadata document that explains the various fields, their usefulness and summarises the purpose of the dataset in general. These documents will be stored along with the dataset and are accessible in the same manner as the dataset (e.g. online, or download). Examples of such accompanying documentation are available for download. If you wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

6.5 Typically, any stored data, materials, artefacts, etc. will need to be cited when accessed and referenced by other researchers. It is useful to provide clear and concise citation information for researchers to access. This can be done via the accompanying documentation (metadata) indicated in 6.4 above. If you wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

Section 7: Resourcing of Data Management

7.1 In the event that this is an individual project or piece of research, your own name should be listed. Include any other staff or assistants are to be involved in the project as well. It may be necessary to include staff from other departments of the University. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

7.2 Funding strategies are often outlined by funders and will include a data management aspect. The costs of any materials, equipment and specialist knowledge will need to be factored to arrive at a reasonable estimate. Include any materials or equipment that will be funded by the University and/or you. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

7.3 As in 7.2 above, funding strategies are often outlined by funders and will include a data management aspect. Typically the University will support on-going research projects, and

assist in facilitating post project maintenance and/or presence of outputs. However, this needs to be confirmed to ensure that the service will be available in the form that is required. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

Section 8: Review of Data Management process

8.1 Funders will need to be informed about how the data management process will be implemented. Provide specific information about how you intend to follow through with the commitments and processes that have been discussed in the rest of this document. Typically, regular reviews, reports and assessments of progress will suffice, but some funders may require specific means of identifying adherence to the plan. If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

8.2 Based on 8.1 above, list those who will be carrying out the reviews and subsequent reports or processes necessary to ensure the successful implementation and completion of the data management plan. Typically, in the event of smaller research projects or individual research, the project Supervisor will fill this role. In the event of PhD research, this role will be carried out by the PhD Supervisor(s). If you are uncertain, or wish to explore this avenue further, the first line of contact is the project Supervisor, or the Head of Research.

Section 9: Statements and Personnel Details

9.1 The Statement of Agreement is necessary to clarify the areas of responsibility and work that will be carried out by the various researchers engaged in the project. This information is vital for funding bodies that will require these details.

9.2 As in 9.1 above, the Expertise of Researchers is necessary to clarify the areas of responsibility and work that will be carried out by the various researchers engaged in the project. This information is vital for funding bodies that will require these details in the form of a brief résumé for each researcher.

Section 10: Appendices

10.1 Assistance with completing the Plan; follow the instructions to obtain help specific to each section.

10.2 Follow the guidance for each specific section as necessary.

10.3 This list of Relevant Contacts will be reviewed and altered regularly.

10.3 Relevant Contacts

The following list of contacts will be regularly revised as appropriate.

History Head of Research:	Peter Wilson Grant Professor of History Phone: +44 (0)1482 465382 Email: p.h.wilson@hull.ac.uk
Library Services:	Brynmor Jones Library Reception Desk Phone: +44 (0)1482 466581 Brynmor Jones Library Service Desk Phone: +44 (0)1482 465250
IT Services:	IT Helpdesk Phone: +44 (0)1482 462010 E-mail: help@hull.ac.uk
Head of Department:	David Starkey Reader Maritime History Phone: +44 (0)1482 305114 Email: d.j.starkey@hull.ac.uk
Document Author:	John Nicholls Research Fellow Phone: +44 (0)1482 305119 Email: j.nicholls@hull.ac.uk

Appendix 2: Example MODS record for a dataset in Hull's repository

<mods xmlns="http://www.loc.gov/mods/v3" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-4.xsd">

<titleInfo>

<title>HMAP Dataset 01: SE Australian Trawl Fishery I: Fish (various) landings and fishing effort, South East Australia, 1918-1923</title>

</titleInfo>

<name type="personal">

<namePart>Klaer, Neil</namePart>

<role>

<roleTerm type="text">Author</roleTerm>

</role>

</name>

<typeOfResource>dataset</typeOfResource>

<genre>Dataset</genre>

<originInfo>

<publisher>Barnard, Michaela G,; Nicholls, J. H.</publisher>

</originInfo>

<language>

<languageTerm type="text">English</languageTerm>

<languageTerm type="code" authority="iso639-2b">eng</languageTerm>

</language>

<physicalDescription>

<internetMediaType>application/pdf</internetMediaType>

<internetMediaType>text/plain</internetMediaType>

<internetMediaType>application/ms-word</internetMediaType>

<internetMediaType>application/zip</internetMediaType>

<internetMediaType>application/ms-access</internetMediaType>

<digitalOrigin>born digital</digitalOrigin>

<extent>Database: 36102 records</extent>

</physicalDescription>

<abstract>On 28 November 1991, the New South Wales Department of Agriculture and Fisheries provided historical data relating to the South East Fishery (SEF) to the Bureau of Resource Sciences (BRS) in accordance with the Fishing Industry Research and Development Corporation (FIRDC)-funded 'historic SET catch data' project. The data were originally collected by CSIRO and handed over to NSW Fisheries when CSIRO ceased work in the 1960s on what was then known as the 'South East Trawl Fishery'. These data cover the periods 1918-1923, 1937-1943 and 1952-1957. No documentation exists for these data except for a simple description of the data field names. This document describes the processing carried out, and the assumptions made, to convert the data into a format suitable for inclusion in the Australian Fishing Zone Information System (AFZIS). This format, in turn, was adapted to render the data compatible with the HMAP/OBIS schema.

The map below gives an indication of the extent of the Southeast Australian shelf and slope; the 'view as map' link in the download panel at the right will show a much more detailed representation. The kml file download, when used with Google Earth, will render the extent of the Southeast Australian shelf and slope in detail.

</abstract>

<subject authority="UoH">

<topic>Fisheries statistics</topic><topic>History of marine animal populations</topic>

<topic>Trawling</topic>

<topic>Overfishing</topic>

<topic>Australian fishing industry</topic>

<topic>Southeast Australian shelf and slope</topic>

<cartographics/>

<temporal>1918-1923</temporal>

<geographic>Southeast Australian shelf and slope</geographic>

</subject>

<identifier type="fedora">hull:1934</identifier>

<relatedItem ID="publishRef" displayLabel="Publish reference"/>

<location>

<url usage="primary display" access="object in context">http://hydra.hull.ac.uk/resources/hull:1934</url>

<url access="raw object">http://hydra.hull.ac.uk/assets/hull:1934/content</url>

</location>

<accessCondition type="useAndReproduction">Creative Commons Licence: Attribution-Noncommercial-Share Alike 2.0 UK: England and Wales. See: http://creativecommons.org/licenses/by-nc-sa/2.0/uk/</accessCondition>

<recordInfo>

<recordContentSource>The University of Hull</recordContentSource>

<recordCreationDate encoding="w3cdtf">2010-08-23</recordCreationDate>

<recordChangeDate encoding="w3cdtf">2012-04-12</recordChangeDate>

<languageOfCataloging>

<languageTerm authority="iso639-2b">eng</languageTerm>

</languageOfCataloging>

</recordInfo>

<relatedItem ID="relatedMaterials">

<location>

<url usage="primary display" access="object in context">HMAP website: www.hull.ac.uk/hmap</url>

</location>

</relatedItem>

<note type="citation">The dataset: please cite as follows: N. Klaer ed. ‘South East Australian Trawl Records, 1918-1923’ in M.G Barnard & J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap)</note><note type="citation">Supporting documentation: please cite as follows: N. Klaer, ‘HMAP Dataset 1: SE Australian Trawl Fishery I, Supporting Documentation’, in M.G Barnard & J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap)</note>

<note type="software">Microsoft Access 2000</note>

<subject ID="location" displayLabel="Southeast Australian shelf and slope (indicative only)">

<cartographics>

<coordinates>138.8621151800028,-35.7533580442386,0 136.6461482736363,-36.73520354575968,0 131.7719329975087,-36.16495120493379,0 140.3413995550836,-40.96033731385748,0 142.5758637433944,-44.6983020087938,0 145.3759003340833,-45.62887457324322,0 147.5626744361646,-45.65300136395928,0 149.8194498326327,-45.11078810578129,0 150.7185692705386,-44.41178446679641,0 152.0219507338506,-40.68144188859979,0 156.0664754616064,-36.28512532901873,0 150.2602181636284,-37.06351215419577,0 149.8805291863619,-37.91142643198369,0 148.3587603689652,-38.03428863942413,0 146.517427705729,-39.32254551701976,0 144.5341094564015,-38.51879881274845,0 143.6056497071093,-39.07654591344959,0 139.4931958433867,-37.31827241285865,0 138.8621151800028,-35.7533580442386,0</coordinates>

</cartographics>

<topic>Polygon</topic>

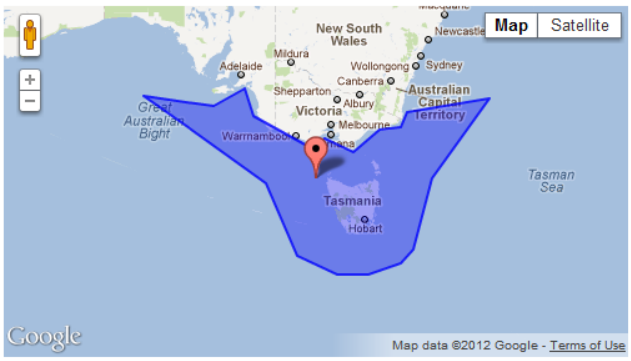
</subject>
</mods>

Appendix 3: Example splash page for a dataset in Hull's repository

« Back to Results Switch to edit view 1 of 19 Next »

Dataset

HMAP Dataset 01: SE Australian Trawl Fishery I: Fish (various) landings and fishing effort, South East Australia, 1918-1923

Person	Klaer, Neil (Author)
Subjects	Fisheries statistics; History of marine animal populations; Trawling; Overfishing; Australian fishing industry; Southeast Australian shelf and slope
Rights	Creative Commons Licence: Attribution-Noncommercial-Share Alike 2.0 UK: England and Wales. See: http://creativecommons.org/licenses/by-nc-sa/2.0/uk/
Citations	The dataset: please cite as follows: N. Klaer ed. 'South East Australian Trawl Records, 1918-1923' in M.G Barnard & J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap); Supporting documentation: please cite as follows: N. Klaer, 'HMAP Dataset 1: SE Australian Trawl Fishery I, Supporting Documentation', in M.G Barnard & J.H Nicholls (comp.) HMAP Data Pages (www.hull.ac.uk/hmap)
Description	On 28 November 1991, the New South Wales Department of Agriculture and Fisheries provided historical data relating to the South East Fishery (SEF) to the Bureau of Resource Sciences (BRS) in accordance with the Fishing Industry Research and Development Corporation (FIRDC)-funded 'historic SET catch data' project. The data were originally collected by CSIRO and handed over to NSW Fisheries when CSIRO ceased work in the 1960s on what was then known as the 'South East Trawl Fishery'. These data cover the periods 1918-1923, 1937-1943 and 1952-1957. No documentation exists for these data except for a simple description of the data field names. This document describes the processing carried out, and the assumptions made, to convert the data into a format suitable for inclusion in the Australian Fishing Zone Information System (AFZIS). This format, in turn, was adapted to render the data compatible with the HMAP/OBIS schema. The map below gives an indication of the extent of the Southeast Australian shelf and slope; the 'view as map' link in the download panel at the right will show a much more detailed representation. The kml file download, when used with Google Earth, will render the extent of the Southeast Australian shelf and slope in detail.
Coverage	Southeast Australian shelf and slope
Temporal	1918-1923
Geo-data	
Related web materials	HMAP website: www.hull.ac.uk/hmap
Extent	Database: 36102 records
Software	Microsoft Access 2000
Publisher	Barnard, Michaela G,; Nicholls, J. H.
Identifier	hull:1934
Language	English
Genre	Dataset

▸ Show additional information

Downloads

- Database - ASCII format (zip)
- Database - Access 2000 format (zip)
- Database - csv format (zip)
- Documentation - PDF format (pdf)
- Documentation - text format (txt)
- Documentation - Word (.doc) format (doc)
- Southeast Australian shelf and slope - kml file (141 KB vnd.google-earth.kml+xml) View as map

QR code link to this page

Appendix 4: Question outline for the initial interviews

Note that, in order to save space, the areas available for answers on this form have been compressed from the original document

History Data Management Plan

Name of interviewee:	
Name(s) of Interviewers:	
Date:	

Which areas of your research do you feel that we may be interested in?
What sort of data does your research give rise to?
What are you doing with your data?
Is your data safe/secure?

How far along is your data in the process of finalisation?
What are the current prospects for your data?
What is being done about this?
What do you see as the intention for your data:

Public access	Closed access	Secure	Curate-able	Other:

What is the Life Stage of your data: (Initial state, raw data, project driven, complete, evolving, etc.)
What is the format, appearance, structure of your data?
Is the data currently available? If yes, how?

Is your data currently being housed in a repository?			
History Data Centre	eDocs	Hydra	Other:
If yes, how appropriate, useful, effective do find this?			
If no, do you want/need to have your data housed? Any ideas about this?			
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.			