

Data Management Plan

example:

**Arts and Humanities Research
Council**



The Professional Career and Output of Trevor Jones

Project Stage: Application

RCUK Research Councils: Arts and Humanities Research Council

Lead organisation: University of Leeds

Project dates: 1 September 2013 to 31 August 2016

Budget: £540,032.00

This DMP, made public with the kind permission of the PI Andrea Hilonutz, represents a real example of a funded proposal from the University of Leeds that forms part of a wider suite of documentation submitted to the research funder. The contents of this DMP should not be directly replicated as a DMP is a document unique to your research proposal only. Please take advice from your Faculty IT Manager or the Research Data Leeds Team (researchdata@leeds.ac.uk) on your requirements prior to making a DMP submission to your research funder.

1 Project Management of technical aspects. You should indicate how you will manage the technical aspects of the project to ensure its timely and successful completion, and clearly state the electronic output(s). Please note that the Case for Support should outline how the project as a whole will be managed. In particular you should address the following:

Test text (TB).

1.a Management and reporting structure

DCC 7.1: Outline the staff/organisational roles and responsibilities for data management

The RA will be involved in the day-to-day technical aspects of the project, though any significant issues will be brought to the PI or CI for discussion as a matter of course, with the PI taking overall responsibility for management of the project's technical aspects. The RA will be the principal liaison with industry specialists, Library personnel, the repository and RoaDMaP teams, and the website design team in particular. The RA will report directly to the PI as the manager of the overall research project.

1.b Project timetable

Most of the technical work in year 1 will relate to the digitising of film-music materials, either in-house or through external companies as appropriate. In-house digitisation of paper-based materials will require a computer, scanner and digital camera, while materials processed externally (whether audio, video or textual) will be digitised by experts in the field. All materials will be digitised in non-lossy formats and in line with good practice guidelines set out by Library and RoaDMaP personnel. Year 1 will see the digitisation of all analogue sound reels and the majority of the related paperwork, with the digital files delivered to the RA as they become available. Basic metadata relating to the materials will be entered into the repository with the digital files.

1.c Project deliverables

DCC 2.1: Give a short overview description of the data being generated or reused in this research

The main project research will be supported by an integrated digital repository. Accessed via a bespoke web interface (hosted by the University of Leeds), the repository will be fully-searchable and will contain all of the digitised film-music materials, while also providing secure storage and suitable redundancy for materials to be preserved and accessed long-term. The electronic resource is of fundamental importance for the project research, which cannot be carried out without it. The University's institutional repository system is managed by a designated team of specialists and includes all necessary security protocols and redundancy to ensure no data are lost even if there is a catastrophic system failure within the repository.

1.d Monitoring process

DCC : When will adherence to this data management plan be checked or demonstrated?

The repository will be accessed via the University of Leeds website and will observe all regulations imposed by the institution in terms of good practice and accessibility of websites. On a more immediate level, the PI and particularly the CI will be responsible for monitoring the standards of work of the RA on the resource, periodically inspecting the collected metadata and testing the robustness of the system and web interface to ensure they function as intended. The Special Collections staff at the University of Leeds's Brotherton Library will assess and monitor the presentation of the digital materials from a bibliographic viewpoint, and RoaDMaP personnel will advise on the required metadata for repository items. Additionally, the CI is a member of the RoaDMaP team and will ensure this project meets standards and good practice developed by that JISC-funded project.

DCC : Who will do this?

See above

2 Data development methods. In describing data resource development methods you should demonstrate your knowledge and application of best practice. You may wish to

focus on some or all of the following issues.

2.a Content Selection

DCC 2.2.1: Have you reviewed existing data, in your own institution and from third parties, to confirm that new data creation is necessary?

Yes

DCC 2.2.2: What existing datasets could you use or build upon?

None

DCC 2.2.3: Describe any access issues pertaining to the pertinent, existing data

N/A

DCC 2.3.2: Describe the process by which you will capture/create new data

The content will be a limited dataset comprising the film-music materials held in the Trevor Jones Archive, and as such the actual selection of items for digitisation and placement in the repository is uncomplicated. Paper-based resources will be scanned or digitally photographed as appropriate; audio and video materials will be sent for digitisation by an external company following normal University of Leeds procedures, including obtaining an appropriate number of competitive quotes for the work. Metadata will be collected for all items in the collection to ensure the repository can be searched satisfactorily and that retrieval of items for use in the research is efficient and comprehensive. Standards for metadata and the presentation of digital materials will be agreed with experts from the University of Leeds Library Special Collections and central IT staff, and the RoaDMaP project team will assist in ensuring research materials are stored securely in line with best practice.

2.b Please outline and justify the data / file formats that you propose to use

DCC 2.3.3: Which file formats will you use, and why?

All textual items will be scanned or photographed as lossless, uncompressed, colour TIF or RAW files at 300dpi to ensure preservation-quality digital images. Low-resolution (72dpi - screen resolution) copies of all images will be made for use as thumbnails to reduce load times for web pages when the project team are browsing or searching the materials, with the high-resolution images called by the system when individual items from the collections are accessed for use. Audio materials will be stored as lossless WAV or FLAC files at 16 bit/44.1k to ensure high-quality sound without excessive file sizes. It is anticipated that video sources will be digitised in QuickTime format or equivalent, the exact specifications to be decided in consultation with the company that carries out the digitisation and the technical experts at the University of Leeds, to ensure that the resulting digital files are of high-quality.

2.c Documenting the resource

DCC 2.5.1: Are the datasets which you will be capturing/creating self-explanatory, or understandable in isolation?

Yes

DCC 2.5.2: If you answered No to DCC 2.5.1, what contextual details are needed to make the data you capture or collect meaningful?

N/A

2.d Advice sought on planning your proposed project

DCC 1.3.1: Funding body requirements relating to the creation of a data management plan

Completion of the Technical Appendix within the Je-S system.

DCC 1.3.2: Institutional or research group guidelines

Completion of a DMP.

DCC 1.3.3: Other policy-related dependencies

None

2.e Consultation with projects using similar methods

The RA on the AHRC-funded 'Nineteenth-Century String Editions' project (2008-2012, PI: Prof Clive Brown) was consulted regarding the creation of a bespoke web interface for a digital collection, and the nature of the metadata collected for individual digital items on that project. Additionally, he advised that a digital camera would probably be required for the photographing of paper-based objects for which scanning was impractical (such as bound scores), leading to the inclusion of this equipment in the project costs.

3 Infrastructural Support

3.a Describe the hardware, software and relevant technical expertise that is available to you

The repository system is well established and, having been developed out of MIDESS, has been subjected to appropriately rigorous testing and technical refinement. There is a one-off up-front charge per terabyte of storage, which will be maintained long term by the University with appropriate levels of redundancy and security. Furthermore, there are dedicated University personnel with the required technical expertise to manage this aspect of the project. Backup of the data is inherent in the repository (see 3c, below), and will maintain best practice as determined by the outcomes of RoaDMaP. Digitisation carried out by external companies will be undertaken by experts, with companies selected following a competitive tendering process, and technical standards and requirements will be negotiated and agreed in advance of work starting. General technical support will be available through the School of Music and central IT services at Leeds.

3.b Indicate what additional hardware, software and relevant technical expertise, support and training is likely to be needed and how it will be acquired

The hardware required for the in-house digitisation - a computer, sheet-feeder scanner, digital camera and printer - is either already available or will be purchased using a designated part of the research grant. The project team, particularly the RA, will work closely with members of the repository team, ensuring that additional support is available if and when it is required. The web interface will be designed and created by an external company following the model used in the AHRC-funded 'Nineteenth-Century String Editions' project. The project will be supported at an institutional level by the central IT service, the University of Leeds Library, and RoaDMaP, as well as the School of Music and Faculty of Performance, Visual Arts and Communications.

3.c Describe the backup procedures that your project will use to safeguard your electronic resource during its development.

DCC 5.2.1: How will you back-up the data during the project's lifetime?

Electronic data will be stored on the University of Leeds SAN (Storage Area Network), which comprises enterprise level file servers in physically secure data centres with appropriate fire suppression equipment. Snapshots are taken every day at 10pm (and accessible for 1 month). A second level of snapshots is taken every month and are kept for 11 months. Snapshots are user recoverable from the desktop.

An incremental copy to backup tape is taken every night (and kept for 28 days) and a full copy is taken every month. Every quarter, the full dump tapes are moved to a long term storage facility where they are kept for 12 months.

Tapes are initially stored in on-campus fireproof safes and then moved to off-campus secure locations.

Once the data has been manipulated and no further changes are required, it will be moved to an area of the SAN with a less frequent back-up policy.

4 Data preservation and sustainability

4.a You should demonstrate that you have sought advice on any issues which apply to the resource and its preservation

DCC 6.1: What is the long-term strategy for maintaining, curating and archiving the data?

Advice has been taken from members of the repository and Library teams regarding the collection of metadata to ensure that preserved items can be found within the repository, and Library, IT and RoaDMaP personnel and external technical experts have been consulted in order to ascertain appropriate file formats and resolutions for preserving the digital materials long-term without loss of detail. One of the project research questions relates to the application of the digital repository framework in the industrial film-making process, and the advice and guidance of the repository team has been invaluable in developing an understanding of the ways in which digital objects can be preserved and used.

Additionally the external company selected to carry out the digitisation of audio and visual materials will advise on appropriate formats and standards to ensure the digital materials are properly preserved and sustained through and beyond the life of the project. It is anticipated that few (if any) proprietary file formats will be used, to facilitate long-term preservation and future use.

4.b You should indicate what plans you have to preserve the data and make it available in an accessible repository beyond the end of the grant

All digital data will be held in a secure repository, and accessed via a bespoke web interface. The repository has built-in redundancy and rigorous back-up systems to ensure that all data are preserved properly and held securely. Storage of and access to materials will be maintained by the University of Leeds beyond the end of the grant, with the ability to retrieve and use all items in the collections. In the extremely unlikely event of a catastrophic system failure of the repository, the University's central IT team

would supervise the restoration of the data from the most recent backup. The research materials would be available to scholars and researchers beyond the end of the grant period for use in research and teaching, and additional metadata generated from ongoing research would be appended to digital items as appropriate.

4.c You should indicate what plans you have to ensure that the electronic output will become a sustainable resource

DCC 7.3: How will longer-term data management activities be funded after the project ends?

The repository system is fully supported by the University of Leeds, so access to data will be maintained through future advances in technology, and the systems will be updated as web-standards and environments change. The University will manage and preserve the repository beyond the grant period, making it a long-term sustainable research resource. Adherence to the recommendations of the JISC-funded RoDMaP project will ensure data are stored and managed in line with best practice, ensuring the sustainability of the resource in the long term.

Owing to the scarcity of the sorts of materials to be held in the repository (as outlined in the case for support) it is unclear what time-span is required for storage, but the technology employed will ensure that all materials remain accessible for at least ten years beyond the end of the project. Additional items could be added to the repository as and when Jones undertakes further scoring projects, and additional metadata generated from ongoing research would also be added as appropriate.

5 Access

5.1 You should indicate what plans you have to make the resource accessible to the potential audience(s) you have identified

DCC 4.1.1: Are you under obligation or do you have plans to share all or part of the data you create/capture?

Yes

DCC 4.1.3: If you answered Yes to DCC 4.1.1, How will you make the data available?

The purpose of the digital repository is to make the archival film-music materials available to the project team for the purposes of answering the research questions, and it is not intended that there will be public access to the repository. The digital materials will also be used in teaching at the University of Leeds, and other scholars (from Leeds and elsewhere) may be given access to it for use in research and teaching, in accordance with the agreement held between the University of Leeds and Trevor Jones. Access will be limited using a password system. The original artefacts (e.g. the actual sound reels and their boxes, etc.) will be preserved in Library stores at Leeds, though all data associated with the artefacts would be stored digitally in the repository and it is unlikely that scholars would have any need to access the original items.

DCC 4.1.4: If you answered Yes to DCC 4.1.1, When will you make the data available?

The data may be made available to scholars outside the project team once the system has been set up and testing has been completed.

DCC 4.1.5: If you answered Yes to DCC 4.1.1, What is the process for gaining access to the data?

Scholars will apply for access and be allocated a username and password. There may be a charge for access, though this is yet to be decided.

6 Copyright and intellectual property issues

6.1 You should demonstrate that you have sought advice on and addressed all copyright and rights management issues that apply to the resource

DCC 3.2.1: Will the dataset(s) be covered by copyright or the Database Right? If so give details in DCC 3.2.2, below.

Yes

DCC 3.2.2: If you answered Yes to DCC 3.2.1, Who owns the copyright and other Intellectual Property?

Given copyright limitations and the scope of the agreement currently in place between Trevor Jones and the University of Leeds, the digital materials will only be made available to scholars for use in research and teaching. Advice has been sought from the University of Leeds' legal team and experts in the University's Brotherton Library who have significant experience in the area of copyright and intellectual property. All copyright materials will be stored electronically at extremely high levels of security to further ensure that they are accessed and used appropriately. Discussion with composer is ongoing with regard to making some of

the materials (or samples of the materials) available on an open-access basis, and it is anticipated that these discussions will continue as the project progresses.

- 7 FURTHER NOTES. In addition you should note requirements for deposit outlined in the AHRC Annex to the Terms and Conditions of Research Council Grants. Gantt charts or other relevant graphical information should be included as an attachment to the proposal if this is considered necessary; again reference should be made to this within the appendix.**

None
