What counts as data for Arts and Humanities subjects?

The content of a Data Management Plan applies to data created by a research project. Data qualifies as the research collected that directly informs a research output, such as a published article. This can include researcher notes, memos/analysis and commentary, as well as photographs taken in an archive, sound files and traditional data sets, for example. The Data Management Plan is a document that addresses how this information will be stored, managed, and accessed both during and after a project. In this sense, a Data Management Plan describes the information management practices that researchers already have in place.

A Data Management Plan divides into four parts: Summary, Collection, Storage and Sharing. Ethical and Legal Considerations must be considered but are included separately. Each section asks the researcher to identify and address different aspects of how research content is managed over the lifecycle of a project.

Data Summary

In the Data Summary section, please identify the data type(s) used, and the reasons for using these data type(s). Data can be: spreadsheets, drawings, images, videos, surveys, interviews, maps, physical samples, field notes, source material, notes you produce – basically anything you generate/collect as part of your research.

Briefly introduce the types of data the research will create with clear descriptions of the types of data, details of formats and approximate dataset sizes where appropriate. The UK Data Service Recommended formats can be found here.
Explain why a specific data type and format is being used and how it will contribute to the project/answer the research question.

Data Collection

The Data Collection section asks you to describe how you will collect and organise data to make the information easier to find, understand and reuse. Methodologies that address version control are important to consider, particularly if research content is generated and accessed by several members of a team. How will consistency and quality of data be managed and maintained?

Describe the methods used to collect and organize your research data. An example statement might be: “The image files will be stored in thematically organised folders, by date. They will be tagged with keywords to allow cross-searches.” Please provide information on any intermediate data type produced (if any).
Also explain which methodology will produce which data type.
Where methodologies use specialist techniques, equipment or processes, clearly describe how the research will be supported by both the project’s research team and the institution’s data support teams.
Describe existing datasets that will be used, if any. If the project will use existing datasets, how is the data licensed?

If any personal or sensitive data (e.g. in relation to human participants, or any culturally, environmentally or commercially sensitive information) is being collected as part of the project, then this should be described. If human participant data is part of research then describe how consent will be collected.
Provide detail on how the research team's capacity is appropriate for the data/digital aspects of the work.
Explain how you will maintain consistency and quality. Strategies may include a regular review of processed research to ensure standards are being met; organising and archiving email correspondence using proper controls; bracketing photographs.

Data Storage

The section on Data Storage should describe both short-term and long-term data storage plans. Short-term data storage includes how research is backed up during the project. Long-term storage includes depositing research data in a repository such as that at the UK Data Service or the Apollo Repository at Cambridge University Library. The University's Records Statement and Retention Schedule acts as a useful guide for best practice standards.

Describe and justify where each type of data will be stored and backed up during your research. All data should be backed up in at least three places: laptop/desktop, external drive and the cloud.
Identify who is responsible for managing your storage and backups, and how often backups will be carried out (method, location, frequency, quantity and server, hard-drive and cloud).
Reference institutional storage policies or pages where necessary: University of Cambridge Research Data Management Policy Framework.
Where data contains sensitive information or is considered high-risk, the research team may want to develop and implement further security procedures. An example statement might be: "We recognize that this data is highly confidential and is critical to [state issues and effected groups]. Therefore a project-specific security policy has been developed in conjunction with the University's Policy on the Ethics of Research Involving Human Participants and Personal Data." An Information Security Risk Assessment (ISRA) assists in determining risk level and security actions.
Where data does not contain sensitive information or is considered low-risk, an example statement might be: "The data will not include personal data relating to human participants. The University's Research Integrity Statement, Good Research Practice Statement and policies and procedures for research ethics will be abided by at all times."
For long-term data storage, provide details about the repository such as the UK Data Service or the Apollo Repository or other appropriate solution for retaining data, and outline the retention period appropriate for the data that is in keeping with consent from participants.
Provide details about the cost of long-term data storage. Costs can vary between repositories.

Data Sharing

The Data Sharing section asks you to consider the value of the data for future users and how and when it will be accessible. This includes details about accompanying documentation that explains how the data is organized. While UKRI funding requires that a project's data be made publicly accessible, it is possible that not all data collected can be stored long-term. For instance, archives can limit how photographs of their collections can be used. Limitations to open access of data must be explained in the data

management plan. Access to sensitive data must adhere to participant consent. Guidance on safeguarding sensitive data can be found here: Research Ethics.

Outline the value of each data type, considering how value is facilitated by data release and/or collaborations.
Describe how the research data is useful to possible future audiences. Any limitation on future use is described and justified.
The point at which data will be released is clearly identified. This might be a date (in relation to the lifetime of the project) or triggered by an event such as the publication of major findings. If there are reasons for delaying the release date, please justify.
Outline where the data will be shared: for instance, in which repository or repositories?
How will different audiences be made aware of this date? Describe discovery routes such as data availability statements in associated papers, statement of availability on the project website, whether the repository is searchable and keywords, etc.
Will the data require future updating and if so outline an appropriate plan that includes frequency of updates, and the point at which updates will no longer be required.
If applicable, the license and charging model for datasets are clearly described and justified. For more information about commonly used licenses, see Data Research Licences.
Reference institutional data sharing policies or pages where necessary: University of Cambridge Research Data Management Policy Framework
Reference institutional statement on open research: University of Cambridge Open Research Position Statement

Ethical and Legal Considerations

Outline any ethical and/or legal considerations in collecting the data (note ethics approval or intention to seek ethics approval if necessary). Also note these considerations in relation to storing and releasing the data; noting anonymisation, etc. of participants in accordance with agreements made with participants. Ethical and legal considerations may also apply to non-personal data that is culturally, environmentally, or commercially sensitive.

If IP is part of the data then describe how permissions will be gained for future usage. For more information about commercialising data, please be in touch with Cambridge Enterprise.
Are there legal, ethical and security issues around the storage (both short and long term) and future use of the data? What solutions or mitigations have been put in place to address these issues?

Further guidance

Cambridge University Library offers Data Management Plan Support and has additional literature and guidance on how to design and implement a data management plan that can be found in the Data Management Guide.

As funders have different data management requirements, the Summary of Data Research Policies by Funder may be helpful as a quick reference.

Examples of data deposited in Apollo

Photographs: https://www.repository.cam.ac.uk/handle/1810/264602

Working documents: https://www.repository.cam.ac.uk/handle/1810/294062

Statistical study of historical material: https://www.repository.cam.ac.uk/handle/1810/274103

Creating data from historical sources: https://www.repository.cam.ac.uk/handle/1810/319574

Field-note book: https://www.repository.cam.ac.uk/handle/1810/268245

Appendix: https://www.repository.cam.ac.uk/handle/1810/318231

Transcription: https://www.repository.cam.ac.uk/handle/1810/244204

Large Scale Project: https://www.repository.cam.ac.uk/handle/1810/253889

This document compiles information made publicly available by University of Cambridge, University of Bristol and The University of Sheffield, as well as from workshops and advice from Data Champions at Cambridge. A list of Data Champions and their areas of expertise can be found here.