
Plan Overview

A Data Management Plan created using DMPonline

Title: University of Hertfordshire Example DMP

Creator: Danny Smith

Principal Investigator: Danny Smith

Data Manager: Danny Smith

Affiliation: University of Hertfordshire

Template: Data Management Plan Template

ORCID ID: 0000-0002-0724-2374

Project abstract:

This example DMP uses the University's Data Management Plan Template to demonstrate the sort of content we expect our authors to produce in their DMPs. Some of the content included in this example is taken from actual DMPs produced by our authors, whilst some has been fabricated expressly for the purpose of this example, hence the plan may discuss medical data at one point and astrophysical data the next. We have done this so as to provide an example that is useful to researchers in as many disciplines as possible.

ID: 68478

Last modified: 17-12-2020

Grant number / URL: [e.g. NIHR12345]

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

University of Hertfordshire Example DMP

Contact Information

Name:

E-mail address:

School:

Date: [date of first completion then subsequent dates of revisions e.g. 13/10/2020 // 20/12/2020]

Version: [v1 if this is the first version of the DMP produced, v2 if second, etc.]

DMP Author: Joe Bloggs

E-mail: e.g. ProjectTeam@herts.ac.uk

School: School of Life and Medical Sciences

Date: 17/12/2020

Version: v1

Project Information

Project name:

Funder:

Grant number:

Project summary: [Short description of the project's fundamental aims and purpose]

Project name: University of Hertfordshire DMP Example Project

Funder: e.g. NIHR

Grant number: e.g. NIHR12345

Project summary: This example DMP uses the University's Data Management Plan Template to demonstrate the sort of content we expect our authors to produce in their DMPs. Some of the content included in this example is taken from actual DMPs produced by our authors, whilst some has been fabricated expressly for the purpose of this example, hence the plan may discuss medical data at one point and astrophysical data the next. We have done this so as to provide an example that is useful to researchers in as many disciplines as possible.

Data Overview

Data summary: [Short description of the data being generated or reused in this research]

Pre-existing data: [List any relevant pre-existing data sources to be used]

Data capture: [Methods of data collection; will any community standards/guidelines be adhered to?]

Type(s): [e.g. quantitative/qualitative, survey responses, video footage, models, images, medical records, etc.]

Format(s): [CSV, PDF, MP4, JPG, TIFF, TXT, DOCX etc. If using proprietary or unusual formats, state why]

Expected quantity: [e.g. 4 spreadsheets, 64 images etc.]

Expected size: [1GB, 500MB, 2TB etc.]

Data summary: The observational projects described in the proposal will generate images, spectra and other forms of raw telescope data (radio measurement sets, optical images and spectra). Data reduction will entail producing calibrated images, spectra and data cubes usable for science from these raw data, and in some cases generating other datasets such as catalogues.

Pre-existing data: Because of the novelty of Mobility as a Service (MaaS), there is a scarcity of qualitative and quantitative data on usage of MaaS collected to document consumer individual initiatives and practices, because real examples of MaaS are few and far between, most being demonstrator projects. A review of the academic literature and reports has revealed that data exists collected on attitudes but there are no data based on social sciences theories. There is a deficiency of rich qualitative data as well as quantitative data.

Data capture: Data generated in this project will be handled in accordance with JIC's Quality Assurance Policy, which conforms to internationally recognised best-practice. Step by step good practice for image-based profiling has been produced in the context of the cytomining 'hackathon', in which a number of lead institutions came together to discuss and share their expertise in morphological profiling. References: Caicedo, J., Cooper, S., Heigwer, F. et al. Nat Methods 14, 849-863 (2017)

Type(s): Quantitative data: movement trajectories, and muscle activity from experiments. Quantitative and qualitative data: on activities of daily living generated from questionnaires, written diaries and also audio recordings of interviews with participants.

Format(s), Quantity and Size:

- Audio recordings - circa 200 between 1.5 hours and 2.5 hours in uncompressed Waveform Audio File format (.wav)
- Written diaries - circa 400 single page OpenDocument Text files (.odt)
- Motor control trajectory data will be recorded via a Matlab program as spreadsheets in comma-separated values files (.csv)
- Based previous experience we expect to collect 5600 files of data with a total file size of 40 GB.

Instead of typing out the details of your data formats, quantity, and sizes, it can be better to display this information in a table, e.g. below.

Data Types	Format	Data Volume	Quantity	Open software*
Python custom scripts for data processing	.py	2-5 KB/file	20	Text editor
Microscopy image stacks (INCA 6000)	.xdce	30 GB/file	60 (~2 TB)	CellProfiler Image J
Cell by cell processed data, expression data	.csv	3-4000 KB/file	60	LibreOffice
Flow cytometry	.fcs	1500 KB	30	FlowingSoftware
Other biological results (e.g. fluorescence, wb, ELISA)	mixed	50-4000 KB/file	200	LibreOffice
Model (description and protocols)	.pdf .txt	20-1000 KB	15	LibreOffice
Outreach evaluation data	.pdf .csv	20-1000 KB	60	LibreOffice

*All data formats will be non-proprietary allowing all users to access the data via open source software platforms. Full details for each format will be available in a ReadMe file which will accompany the data.

Data Curation

Data integrity: [List file naming conventions, data validation measures, peer review of data and other measures being taken to ensure the quality of the data]

Metadata: [What metadata will you produce about your data? E.g. documenting methods used to generate data, recording instrument data along with data itself, ReadMe files, audit trails etc. What format will your metadata be in? See here for help with [metadata](#)]

Data integrity: We plan to use questionnaires developed by others, such as the RoSaS or Godspeed, that have been peer-reviewed and validated, and are considered standard in the HRI community. All movement trials will be repeated multiple times to ensure consistency and quality of data collection as is the norm for motor control experiments and consistent with our previous work in this field (Yousif et al, J Neurophysiol 2012; Yousif et al., Exp Brain Res, 2015; Hammerbeck et al., Neurorehab & Neural Repair, 2017). Transcripts of interviews will be inspected by the Principal Investigator (PI) to ensure accuracy. The spreadsheets including quantitative data will be inspected by co-investigators for duplication.

Metadata: Project-level documentation will be stored alongside the data and continually updated during the life of the project. Data/Object-level documentation will be recorded by the instruments used in data collection and in most cases embedded within the

data object itself where possible.

The methods used to generate the data will be documented and the metadata for each experiment will be stored with a .txt ReadMe file which will contain the specific codes used for the trial types and all the relevant variables which appear in the movement data files. These files will include coding information of the data including memos, which means other researchers will be able to interpret the data.

Security and Storage

Anonymisation measures: [If you will be anonymising the data describe how]

Data storage: [Where will the working data be stored during the project? How will it be backed-up, where, and how often? How will working data be shared during the project? See [here](#) for help with data storage.]

Data security: [How will you manage access restrictions and data security during the project's lifetime? What are the main risks to your data and what will you do to mitigate these risks (e.g. loss/theft/damage)? How will you protect data in transit?]

Anonymisation: Transcripts of interviews and fieldnotes will be anonymised in accordance with best practice as per the Information Commissioner's Office's *Anonymisation Code of Practice*.

Data Storage: All data will be stored on the University's dedicated and secure research storage service (known as R drive). This networked storage is user access controlled, backed up daily, and sits in the University's secure network; only researchers involved in the project will have access to the R drive. A logical file naming system will be used when storing data and all data will be encrypted before any transfer takes place. The data retention period of the funder will be followed or the University's 10-year period.

Data Security: If any consent forms are completed on paper and posted back to the research team, they will be scanned and stored electronically and the paper forms shredded once scanning and storage on R Drive has been completed. The University's secure, firewalled network meets all the necessary security standards (in October 2020 we received a Cyber Essentials Certificate of Assurance from the National Cyber Security Centre) and access is limited via user access control.

All data that is transported between the field, the collaborators and the host institution will be transferred via the University's [Exchange File](#) platform, which encrypts all documents/data during transfer.

Data Archiving and Preservation

Archiving: [Where will you archive your data, in [a repository](#)? Which one? Will you archive all of your data, or only some? If disposing of your data, explain why and how]

Preservation: [What preservation methods/services will be used to secure the data for future use? For how long will data be preserved? See [here](#) for info on preservation]

Depending on your plans for Data Sharing, you could combine this section with that below.

Archiving: At the end of the project we will assess the data to determine which data is worthy of preservation and which can be discarded, and ensure that accompanying metadata and documentation are in place. Once this has been done, data will be deposited to our institutional repository UHRA.

Preservation: UHRA is supported by the Arkivum cloud-based preservation service and the data retention period of the funder will be followed or the University's standard 10-year period.

Data Sharing

Availability: [Who may be interested in reusing your data and for what purposes? Will you make your data openly available? All of it or just a sample? When will you make it available? If you will place an embargo or other restrictions on access, detail them here. Explain any restrictions to data sharing e.g. for ethical or legal reasons. See [here](#) for info]

on data sharing]

Location of dataset(s): [Where will you share your completed data/materials? In a repository? Or on request?]

Licence: [If openly available, how open will your data be? What [licence](#) will you attach to it? See here for help]

Identifier(s): [Will your dataset be given a DOI or similar identifier? How will you link your publications to your dataset(s)? See our guidance on [data access statements](#)]

You could combine the section above, Data Archiving and Preservation, with this section if you wish.

Availability: The project team will have exclusive use of the data for a period of time at the end of the project until publication of results, at which point anonymised data will be made openly available.

Location of dataset(s): Data will be shared via UHRA which is the publicly accessible institutional repository of the University of Hertfordshire, registered with the Registry of Research Data Repositories (<https://re3data.org>). Metadata is provided in Dublin Core format and is harvested by Google Scholar, CORE, BASE, and potentially by any other system with OAI-PMH capabilities.

Licence: We anticipate data being made available under a Creative Commons Attribution Licence (<https://creativecommons.org/licenses/by/4.0/>).

Identifier(s): Datasets will be assigned a DOI to increase discoverability and any publications related to this project data will be linked to the dataset by use of data access statements.

Responsibilities and Resourcing

Roles and responsibilities for data management: [Who has authored this DMP? Who else from the project is responsible for reviewing it? List any other responsibilities in terms of data management]

Review: [When will this DMP be reviewed and by whom?]

Legal obligations: [List any potential [legal issues](#) with your data]

University regulations: [List any relevant [university policies](#) on data management or associated issues]

Roles and responsibilities for data management: The PI assumes ultimate responsibility for data management but Dr. J. Bloggs will be in charge of the quality assurance of the data and Dr. J. Doe of the metadata creation and updating of the Data Management Plan as the project evolves.

Review: A DMP is a living document and will therefore be continuously updated throughout the project lifecycle to reflect the evolving workings of the project. This will be carried out by Dr. J. Doe every three months.

Copyright/IP: Copyright and IP for any data produced will belong to the University of Hertfordshire, unless a contract assigning IP to the funder is required.

University regulations: We will abide by the following University of Hertfordshire policies and regulations.

[Intellectual Property \(CA04\)](#)

[Data Protection Policy and Privacy Statement \(IM08\)](#)

[Data Management Policy \(IM16\)](#)

[Research \(RE03\)](#)