

## Outputs Management Plan examples

Below are several real examples of what we consider to be a good output management plan. The examples are from researchers, who work in a range of scientific areas, at different career stages.

### Example 1 – Senior Researcher – Neuroimaging data sharing

1. What significant outputs will your research generate?

The research will generate substantial human neuroimaging and LFP datasets.

2. When do you intend to share these outputs?

Data acquisition for some aspects of the study will continue until relatively close to the end of the Fellowship extension. Thus, analysis and writing up will definitely continue to the end of the project, and potentially for a further 3-6 months after that. Primary publications from these experiments are therefore likely to take up to a year after the end of the fellowship. Once published, data can be shared immediately. Unpublished data would be shared after 3 years.

3. Where will you make these outputs available?

The human neuroimaging data will be shared via the WIN, which has 'Open Neuroimaging' as one of its research themes during the 5-year Centre grant. As part of this research theme, methods of sharing imaging data, analysis tools and pipelines, will be developed with the aim of sharing data openly by the end of this project. The study data will be shared using these tools. For any data that is ready for sharing prior to the availability of these tools, we will use repositories such as OpenfMRI (<https://openfmri.org>).

The rodent LFP datasets will be made available via the BNDU website, which has a facility for data sharing (<https://data.mrc.ox.ac.uk/>) and is already attracting attention from relevant researchers.

4. How will they be discovered and accessed by others?

Once funding is secured, a data-sharing policy will be included on my group website and highlighted at research talks and poster presentations. WIN's Open Neuroimaging policy will aim to aid data discovery, and will likely to be the most common route for new users for the human neuroimaging. The BNDU data sharing site is already being accessed by researchers from around the world.

5. Are limits on sharing required?

For the human neuroimaging, at the point at which participants consent to the study, they are asked to consent to anonymous data sharing. Since all MRI and OPM data is anonymised at source, there should be no issues relating to identifiability. There will be some delay while the study team publish primary results, although as laid out above, this will be completed as soon as possible. Initially, the PI will be the custodian of the data and make decisions on whether to supply a potential new user with the data (no-one has been refused for previous datasets). Once data has been deposited access will be controlled centrally.

6. How will these outputs be preserved?

All raw human neuroimaging data acquired in Oxford is automatically archived to tape at WIN and kept for a minimum of 10 years.

OPM, LFP and behavioural data, which will be initially stored locally on the PCs on which it was acquired, will be backed up to the WIN servers as soon as possible and will be archived once a week to institutionally-managed servers. Data that has been analysed, or is in the process of analysis, is stored on dedicated disk space at WIN (RAID device) that is backed up regularly and written to digital tape regularly; one set of which is stored off-site.

The specific protocols used to acquire the imaging data need to be associated with the imaging data and will be stored both electronically with the data and potentially in hard copy in lab books. Similarly, to ensure analyses can be replicated, the specific pipelines used will be stored in electronic lab books, along with the version numbers of software packages.

7. Describe any resources that you will need to deliver your outputs management plan

No additional resources will be required.

### **Example 2 – Senior Researcher – Genomic data sharing**

i. What data outputs will your research generate and what data will have value to other researchers?

All data produced by the project will be subject to the data management plan and sharing agreements. The research will generate two kinds of high content data: Viral whole genome sequences (HCMV) obtained from clinical material (anonymised patients); human exome sequences obtained from infected patients and controls. The genomic data generating experiments will be performed by UCL Genomics and partners and will follow the SOPs developed by the centre. All equipment is under service warrantee and regularly maintained and operated by fully trained and supervised personnel.

This will provide valuable information on the diversity of HCMV in clinical viraemias and provide a catalogue candidate genetic germline variants in the patients for other researchers to explore.

ii. When will you share the data?

All viral genomic data will be released on publication, in accordance with the principles of UCL's Open Access policy (<https://www.ucl.ac.uk/library/open-access/publications-policy>). Due to current data protection and ethics restrictions, we will not be permitted to release raw exome sequence data into public repositories (see below for access information)

iii. Where will you make the data available?

All data will be collected to local servers in .fastq and bam format. We will follow accepted community standards for metadata such as MIBBI. Exome data will be processed using an in house pipeline, or commercial alternatives (Ingenuity, Congenica) to generate .vcf files for variant analysis. The viral data sets will be lodged in the public short-read archive at EBI

([www.ncbi.nlm.nih.gov/sra](http://www.ncbi.nlm.nih.gov/sra)), and the finished sequences will be deposited in a public database (e.g. GenBank).

iv. How will other researchers be able to access the data?

The viral genome data will be accessible either through the public databases as outlined above, or via direct request to the participants.

v. Are any limits to data sharing required?

We will adopt best practice for data sharing agreements and be aligned with standard UCL policies on privacy, data protection and handling sensitive data (<http://www.ucl.ac.uk/isd/itforslms/services/handling-sens-data>). We are subject to data protection and ethics permission restrictions on the Exome data. This data is only approved for use for the specific application of CMV research. Therefore, the data will be stored in the UCL SLMS data safe haven which has been certified to the ISO27001 standard and conforms to the NHS IG Toolkit. It will be made available to collaborative researchers in the CMV field by application, subject to agreeing to terms compliant with our ethics restrictions. For this purpose, exome data will be anonymised and shared in line with the informed consent provided to prevent identification of participants.

vi. How will you ensure that key datasets are preserved to ensure their long-term value?

We have in place long term data storage facilities in addition to the public resource (see above).

vii. What resources will you require to deliver your plan?

No additional resources are requested. Data will be stored securely and locally in the UCL Research Data Services facility ([http://www.ucl.ac.uk/isd/staff/research\\_services/researchdata](http://www.ucl.ac.uk/isd/staff/research_services/researchdata)). This ensures that electronic data will be held in two locations, and automatically backed up nightly to tape stored at a remote location. A 6 monthly archive system to hold data offsite for a minimum of 10 years is in development.

### **Example 3 – Clinician – Controlled access to sensitive data**

(i) What data outputs will your research generate and what data will have value to other researchers?

The project will result in new databases created from

- Record linkage of data extracted from electronic health records (NHS Lothian) and secondary, quantitative data obtained from health and education administrative databases (Information Services Division [ISD] Scotland)
- Anonymised, participant level data from existing birth registries, cohorts and clinical studies.

These datasets will be valuable for other epidemiological studies.

(ii) When will you share the data?

To allow time for data collection/linkage/cleaning/documentation, we will consider requests for data sharing 4 years after the grant start and aim to provide data within 4 months of appropriate approvals being obtained.

(iii) Where will you make the data available?

Information about the study will be made available on a Co-OPT consortium website (including point of contact/data-sharing policies/how to request data). Details of the NHS Information Service Division (ISD) databases and some contributing databases, and metadata are already available on the websites of organisations that collect the data.

To ensure due process and aid finding of the dataset I will register the study (<https://clinicaltrials.gov>) and protocol (<https://www.crd.york.ac.uk/PROSPERO/>), and publish the protocol and a cohort profile. Reporting will adhere to the RECORD statement, with the process of NLP data extraction, record linkage, code and statistical analysis fully documented. Data fields and

coding information from NHS Lothian records will be fully annotated, and made freely available to other researchers throughout NHS Lothian.

(iv) How will other researchers be able to access the data? and (v) Are any limits to data sharing required – for example, to either safeguard research participants or to gain appropriate intellectual property protection?

All processes will adhere to incoming General Data Protection Regulation.

Generic Research Database REC approval will be sought for the research use and onward sharing of the Scottish data. Making the new Scottish database available for further use by other researchers will require the agreement of Lothian Health Board and National Health Services Scotland (i.e. the Data Controllers in Common). A proposal will be made to the data controllers via the PBPP to store the dataset within the NHS Scotland National Safe Haven, and to put in place an Access Committee to consider requests for access on their behalf. An alternative may be to place the Scottish dataset under the direct control of the PBPP. If a separate Access Committee is established we would envisage membership to comprise an independent chair, representation from the Co-OPT consortium and lay representatives from the Patient Advisory Group and parent PBPP. Data-sharing agreements will be issued and signed before data are released or analyses performed. External researchers will be bound by data-sharing agreements. The responsibilities of the user and the consortium will be defined in writing as part of the terms of reference of the collaboration, the PBPP and REC approval. Linked and source data from Co-OPT researchers will only be shared with approval of the data controllers, as I am not the custodian of the primary data. If approvals allow, however, we will lodge resulting anonymised datasets within a data sharing repository.

(vi) How will you ensure that key datasets are preserved to ensure their long-term value?

Data will be accessed via NHS Lothian Research Data Linkage Service and the NHS National Services Scotland National Safe Haven (NSS National Safe Haven) <http://www.isdscotland.org/Products-and-Services/EDRIS/Use-of-the-National-Safe-Haven/>, complying with NHS Lothian eHealth Security policy. Access is protected by personal password. Only the applicant and approved researchers will be able to access the data. Completion of training in information governance is a prerequisite of using the Safe Haven.

The specific terms and duration of storage will be agreed in writing as part of the PBPP application. The linked data will be available only during the study period. The IPD meta-analysis dataset will be stored in accordance with Wellcome Trust policies with agreement from the Public Benefit and Privacy Panel (PBPP) and data controllers.

(vii) What resources will you require to deliver your plan?

Website - to ensure project is discoverable/provide contact details for data sharing requests

Data Administrator (20% FTE)- funding is requested for a data administrator to maintain records, adhere to legislative requirements, and facilitate data sharing.

#### **Example 4 – PhD student – Population modelling data sharing**

(i) What data outputs will your research generate and what data will have value to other researchers?

The data generated during the study with utility to other researchers will potentially include a part/all of the following:

1. Demographic data on free-ranging dog populations (overall population size and density, age and sex composition, body condition scores, GPS locations of sightings and boundaries of study sites etc.)
2. Number and locations of potential food sources for these populations, including GPS locations
3. Demographic and dog ownership data of people at the study sites (family size and composition, basic socio-economic indicators, number and kind of dogs owned, nature of ownership, dog health practices and vaccination status etc.)
4. Data on number of dogs vaccinated during mass vaccination campaigns, number of dogs captured and vaccine doses administered, signs of clinical diseases (including rabies), biological samples collected (blood/serum, saliva) etc.
5. Sero-conversion data from follow-up capture and sampling
6. Mathematical model(s), model outputs and results of statistical analyses generated
7. Other data that may be deemed necessary during the course of the study.

(ii) When will you share the data?

Any software or programs developed during the project period will be archived online using services such as Zenodo or Github, such that specific versions of the software are issued with Digital Object Identifiers (DOI) at the time of deposit. These will be released for public use after completion of the PhD. All scientific manuscripts generated from the research will be published on open-access platforms such as Wellcome Open Research so that they are available to other researchers free of cost, and supporting datasets will be made publicly available at the time of publishing either as supporting information or on a separate data repository.

(iii) Where will you make the data available?

Data will be made available either as supporting information to scientific manuscripts at the time of publishing or on a separate data repository such as Zenodo, after being issued with a DOI.

(iv) How will other researchers be able to access the data?

Data that can be shared unconditionally will be made available when scientific manuscripts are published or on online data repositories like Zenodo, such that they are readily accessible to other researchers. Data that cannot be unconditionally shared upon publication owing to confidentiality or data protection requirements will be identified as such and a contact email will be provided in relevant publications for data access enquiries by other researchers.

(v) Are any limits to data sharing required – for example, to either safeguard research participants or to gain appropriate intellectual property protection?

It is expected that demographic data of people at the study sites (family size and composition, basic socio-economic indicators) may contain personally identifiable information and location data. All such data will be anonymised prior to storage on online data repositories, and therefore will be available to be publicly shared at the time of publication of manuscripts.

(vi) How will you ensure that key datasets are preserved to ensure their long-term value?

Datasets will be stored in online repositories managed by the university/centre/department or public repositories such as Zenodo, and each dataset will be issued a unique DOI. A record of all datasets generated, with relevant metadata including the DOI, will be created using the Centre

Datahub platform (website) of the centre and this will facilitate easy identification and retrieval of datasets.

### **Example 5 – Early Career Researcher – Cardiology data and software sharing**

(i) What data outputs will your research generate and what data will have value to other researchers?

This project will generate the following materials

- A software for analysis of clinical electroanatomical mapping (EAM) data
- Associated algorithms for quantification of established and novel electrogram parameters
- A large database of EAM data from ablation procedures
- *In silico* models of cardiac tissue
- Electrogram and optical mapping datasets from *ex vivo* human tissue
- Machine learning networks designed to identify ablation targets for atrial fibrillation

The analysis tool(s) I will create will be a valuable resource for the wider research community. Although in the outlined project I will focus my analysis methods on recurrence of atrial arrhythmias, the software I will create will be useful in many EP research settings. To maximise impact, I will release my software open-source. Software will be deposited in publicly accessible platforms, using guidance as set out by the software sustainability institute (doi:10.5281/zenodo.1327310). Publication of the software (or updates) will occur following robust validation and proof of efficacy as described in 'Details of research project'. To enable researchers to use the software effectively, I will provide example datasets and user guides, further to guidance provided in associated publications.

To maximise availability and impact, I will host the software on GitHub. This will allow for the creation of an immediately accessible live link, which will be used for publications, presentations and websites such as institutional and non-institutional profiles and social media. Concurrently, I will use GitHub integration to deposit software to Zenodo. Zenodo is open-access repository where data is stored in CERN's research data centre. Depositing on Zenodo will provide the software with a unique digital object identifier (doi) and safeguard current and future access. As a further step, to aid in version control during development and pre-release, I will store software versions on the University eData repository (UBIRA) under restricted access until release. These measures will ensure that software created in this project will remain accessible for use and further development beyond the timescales of this project.

By releasing my software open-source, I will freely share the algorithms I develop, including machine learning network designs. To facilitate usability of these developments, I will concurrently provide descriptive material, beyond information that is provided in journal publications including:

- Diagrams of overall software architecture
- Pseudocode
- Illustrative diagrams of key algorithmic steps

These resources will maximise impact of outputs, by explaining key developments without the requirement for detailed understanding of their deployment (i.e. the actual code to implement these approaches).

As well as the software tools I propose, this project will utilise clinical datasets, and generate both *in silico* and pre-clinical datasets for validation purposes. Datasets produced or used for validation purposes will be made freely available in line with publication, again by online repository. I will further submit datasets that are of sufficient interest to the research community for publication on PhysioNet, a community resource of complex physiological signals, or data descriptor journals such as Scientific Data.

Further to online depositing, I will share our analysis tools and additional data I generate (e.g. predictors and mechanisms of AF recurrence post ablation) by publication and attendance of relevant meetings. I will target meetings with a focus on the interface between basic science and clinical cardiology, including Computing in Cardiology, British Society for Cardiovascular Research and European Heart Rhythm Association.

#### **Example 6 – Mid Career Humanities and Social Sciences Researcher – Sharing of transcripts and fieldnotes**

- (i) The data and software outputs your research will generate

Two core types of qualitative data will be produced over the course of the project. First, transcripts from recorded interviews with research participants (overwhelmingly research scientists). Second, fieldnotes based upon extensive research undertaken as part of the proposed multi-sited ethnography.

These data have capacity to be of high value to social science researchers with a range of interests include Medical Sociology; Medical Anthropology; and Science and Technology Studies. The data may also be of interest to members of relevant scientific communities (e.g. Sport Science, various sciences of dementia).

All data sets will be produced in Microsoft Word and standard metadata formats will be used wherever possible.

- (ii) When you intend to share your data and software

Participants will be informed of data sharing commitments via a consent form stating that data will be shared "Subject to any data licensing or other contractual restrictions or data sensitivities that may apply, as appropriate I [the participant] agree for the relevant research data collected, to be stored, and/or shared for future use." The terminology used here is intended to support any data/outputs dissemination not identified at this point.

Data will be made available at the point which a sufficient body of material has been acquired to be of value to other researchers. It is intended that this will coincide with the publication a 'marker paper', published in Wellcome Open Research, which will help to promote the availability of the resource to the communities in question.

- (iii) Where your data and software will be made available

Data sets will be made openly available with a CC-BY license through Figshare, a Wellcome Open Research-approved repository. Figshare has a large bank of social science data and will be easily locatable.

(iv) How your data and software will be accessible to others

All datasets will have a persistent DOI which will be used to identify associated publications and raise awareness of the availability of the data. The DOIs will facilitate searching and Figshare itself is also fully searchable. I will also publish a short article, describing the dataset in Wellcome Open Research in order to help researchers discover, access and reference the resource.

(v) Whether limits to data and software sharing are required

All data will be anonymised before being made available. Some fieldnotes may, by their very nature, be impossible to de-anonymise and these data will be redacted in order to ensure anonymity is respected. Participants (scientists) are themselves producing research and it may be that participants are not able/do not want to share the findings of their research. Participants have the right to withdraw their data from my project and, accordingly, such requests will be respected.

(vi) How datasets and software will be preserved

The storage of data in a well-known and widely used repository like FigShare will ensure that the data produced throughout the project will be preserved and of utility to the research community long after the project has concluded.

(vii) What materials your research will produce and how these will be made available

It is not anticipated that any research materials will be produced over the course of this project.

#### Intellectual property

##### What IP your research will generate

If it not anticipated that the research will generate any data that require intellectual property protection.

##### Required resources

It is free to deposit data with FigShare and it is not anticipated that any further resources will be required.