



Writing a Wellcome Trust Data Management & Sharing Plan

Report Version Control

Version	Date	Author	Change Description
1.3	02 September 2014	Gareth Knight	Revision to Q1, Q2, Q4 and Q6 on basis of feedback from David Carr of Wellcome Trust
1.2	6 August 2014	Gareth Knight	Added worked examples and LSHTM-specific information. Several sections were re-written
1.1	9 June 2014	Gareth Knight	Produced first draft based on similar documents by University of Bristol and University of Exeter

This work is licensed under a Creative Commons Attribution 4.0 International License. <http://creativecommons.org/licenses/by/4.0/>



Executive Summary

The Wellcome Trust expects funded researchers to manage and share research data in a manner that maximizes opportunities for future research and complies with best practice in the relevant subject domain.

This guide provides an overview of requirements established at the application stage, offering suggestions for how these obligations may be met and worked examples that may be applied in practice.

Key Expectations for Data Management & Sharing

The Wellcome Trust has published a considerable amount of information on their expectations for data management and sharing. Key messages that they communicate include the following:

- Applicants that intend to create, capture, or generate data must submit a 'Data Management and Sharing Plan' (of 1000 words or less) as part of their research bid
 - The Data Management & Sharing Plan will be reviewed as an integral part of the application; a poorly written plan can have a negative impact on an otherwise strong application.
- Applicants are encouraged to claim costs for resources necessary to manage and share data; these should be clearly described and justified.
 - Institutional services must not be claimed for twice: Institutional resources that are already claimed in indirect costs must not also be included in direct costs
 - Arrangements for post-project data management and sharing must be made during the life of a project. Costs associated with long-term curation and preservation, such as POSF (Pay Once, Store Forever) storage, must be purchased before the close of the project grant
- Research data should be made available in a form that can be shared and reused by others. If this is not feasible (e.g. due to legal, ethical, contractual, or other obligations), restrictions should be clearly described and justified.
- The Wellcome Trust recognize that data management and sharing practices vary in each domain and support approaches that are reasonable
 - The data repository chosen to handle curation, preservation and sharing activities should be appropriate to the material/content type and subject domain.
 - Data formats chosen should be appropriate to the content type and accessible by the target audience.
- A limited period of exclusive use of data is reasonable to enable analysis and publication. However, the applicant must state a specific time period when the data will be made available for access by others.
- Wellcome Trust-funded publications must include a concise and clear statement that outlines how datasets that underpin the research findings can be accessed. Accompanying data should be assigned a Digital Object Identifier (DOI) where appropriate.

Advice and Guidance

For tailored advice and guidance, please contact the Research Data Management Support Service

- Email: researchdatamanagement@lshtm.ac.uk
- Web: http://www.lshtm.ac.uk/research/researchdataman/rdm_service.html

Criteria for writing a Data Management and Sharing Plan

The Wellcome Trust expects a ‘Data Management and Sharing Plan’ to be submitted for proposals that meet any of the following criteria:

- The dataset has clear scope for wider research use, enabling questions to be addressed beyond those raised in the original research project;
- The dataset is likely to have long-term value for research or other purposes;
- The dataset will form the basis of a ‘Community resource’, as defined by the Fort Lauderdale Principles¹ and Toronto statement², which is likely to have broad utility for reference and use by a user community
- The dataset represents a significant output of the research project;

Illustrative examples suggested by the Wellcome Trust of applications that would require a plan include: large-scale genetic association studies of common diseases; genome-wide or large-scale functional genomics studies in a specific organism; and longitudinal studies of patient and population cohorts.

A plan will not normally be required for studies that are generating small-scale or limited data outputs. However, there is a general expectation that projects of this type will make data available to other researchers on publication, and deposit it in an appropriate data repository in a timely manner where possible.

If you are unsure whether you are required to submit a Data Management and Sharing Plan please contact the Wellcome Trust for advice.

Format of a Data Management and Sharing Plan

The Wellcome Trust does not specify a specific template or format for the Data Management and Sharing Plan. Instead, they encourage applicants to develop a plan that is proportionate to the scale of datasets and its likely value to the research community. However, they recommend that the document addresses the following questions³:

1. What data outputs will your research generate and what data will have value to other researchers?
2. When will you share the data?
3. Where will you make the data available?
4. How will other researchers be able to access the data?
5. Are any limits to data sharing required - for example, to either safeguard research participants or to gain appropriate intellectual property protection?
6. How will you ensure that key datasets are preserved to ensure their long-term value?
7. What resources will you require to deliver your plan?

Your plan should demonstrate how you intend to fulfil your data management and sharing responsibilities, identify any risks or obstacles that are likely to be encountered and the steps you will take to address these challenges. It can also help to justify any funding required to carry out activities.

¹ <http://www.wellcome.ac.uk/About-us/Publications/Reports/Biomedical-science/WTD003208.htm>

² <http://www.nature.com/nature/journal/v461/n7261/full/461168a.html>

³ <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm>

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

You should begin by describing the data that will be created, captured, or generated in the study. Relevant contextual information to mention include how, where, when and by whom it will be produced. This should be followed by a statement outlining which data is suitable for sharing and the purpose for which it may be used.

The Wellcome Trust encourages applicants to make research data available for access and use by the wider community. However, they recognize that not all data is suitable for sharing. If data cannot be made available, you should clearly describe the reason for this. For example, it contains 3rd party IPR, personal information that cannot be anonymised, or will be used as part of a patent application.

When considering the type of outputs that will be created and shared, it is helpful to consider the person who will read your research papers and project final report. What data would a research need to verify and recreate the research findings? What documentation would they need to understand the data? How could they use the data beyond the original research context? Could it be incorporated into a meta-analysis or used to address other research questions, for example?

Example:

Data capture will be performed during months 8-20, at which point four types of data will be collected from the (approximately) 500 study subjects:

[1] Questionnaire: An interviewer-administered CAPI questionnaire will be performed covering medical history, health and socio-demographic circumstances

[2] Clinical: A clinical examination will be conducted to measure blood pressure, anthropometry, 12-lead digital ECG assessed for abnormalities; assessment of physical functions (grip strength, chair rise, walk speed) and cognition.

[3] Biological samples: Blood samples will be taken from each subject and centrifuged within 2 hours. Together with serum and plasma, it will aliquoted into 20 bar-coded cryovials and stored at -70C

[4] Interviews: It has previously been found that participant's partners tend to report higher levels of alcohol consumption and more frequent episodes of hazardous drinking than men themselves. We will interview partners of 1 in 4 of the study participants (selected randomly) in order to estimate the extent of under or over reporting of behaviours.

All 4 data types will be usable for future research in some form, subject to appropriate measures being implemented to protect participant confidentiality. Digital and physical outputs will be made available to bona fide researchers for health-related research, irrespective of their institution (university, charity, government, commercial) or location (UK or elsewhere).

Further Information

- LSHTM: Identify Data to Share
<<http://www.lshtm.ac.uk/research/researchdataman/share/datatoshare.html>>
- UK Data Service: Planning for Sharing
<<http://www.data-archive.ac.uk/create-manage/planning-for-sharing>>
- DCC: How to Appraise and Select Research Data for Curation
<<http://www.dcc.ac.uk/resources/how-guides/appraise-select-data>>

2. When will you share the data?

The Wellcome Trust expects applicants to clearly state the timescales for making data available for access. This does not equate to an expectation that data should be made available immediately: they recognize that data creators have the right to a reasonable (but not unlimited) period of exclusive use for their data.

When deciding timescales for making data available, you may wish to consider several factors:

- *Best practice in your subject domain:* Data sharing practice often vary between research disciplines. Genetics and physics researchers, for example, have established practices for data sharing that emphasise early publication of research results⁴, whereas other domains publish at a later date.
- *Study Type:* For longitudinal studies, it is advisable to adopt a granular approach to data sharing, releasing subsets of data at distinct periods, rather than wait until the end of the project.
- *Resource Type:* Community resources, such as those defined in the Fort Lauderdale Principles⁵ and Toronto statement⁶ that are likely to have broad utility should be made available during the project lifetime, in order to obtain feedback from the user community and refine it as necessary.
- *Publication plans:* The Wellcome Trust expects data that support research findings to be made available at the point of publication, so that it can be validated by others. Therefore, it is advisable to include data sharing as a component of publication plans.

Researchers may use publication moratoria to facilitate limited, pre-publication sharing of data with other researchers, while protecting their right to first publication. Any such restrictions on data use should be reasonable, transparent and in line with established best practice. The RDM Support Service can advise on any formal moratoria you wish to put in place⁷.

Example

Research papers written and published during the funding period will be made available with a subset of the (anonymised) data necessary to verify the research findings, in compliance with the Wellcome Trust's OA Policy.

The study team will make digital data (outputs 1, 2 and 4) available within 6 months of project completion. This embargo period is requested to allow time for additional analysis and further publication of research findings to be performed.

Further Information

- Wellcome Trust: When Will You Share Data
<<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#two>>
- LSHTM: Timescales for Data Sharing
<http://www.lshtm.ac.uk/research/researchdataman/share/when_to_share.html>

⁴ <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm>

⁵ <http://www.wellcome.ac.uk/About-us/Publications/Reports/Biomedical-science/WTD003208.htm>

⁶ <http://www.nature.com/nature/journal/v461/n7261/full/461168a.html>

⁷ http://www.lshtm.ac.uk/research/researchdataman/rdm_service.html

3. Where will you make the data available?

The Wellcome Trust encourages researchers to submit their data to an appropriate repository or archive, which will handle the process of curation, preservation and sharing on their behalf. Similar requirements are stated in the School's Research Data Management Policy⁸, which states that a copy of the data should be offered to an appropriate data repository or enclave following completion of the work and held for a period that meets relevant requirements.

The Wellcome Trust recognise that data sharing practices vary between research disciplines and will change over time. Therefore, they do not mandate the use of specific data repositories or data sharing systems. Instead, they offer suggestions for repositories suitable for specific domains⁹ and allow the Principal Investigator to decide where they will deposit their data.

LSHTM operate several services that may be used for data sharing:

- *Data sharing during the project lifetime:*
 - Teaming (<http://teaming.lshtm.ac.uk/>)
 - SFX (<https://sfx.lshtm.ac.uk/>)
 - Filr (<https://filr.lshtm.ac.uk/>)
- *Post-project data curation & sharing:*
An institutional research data repository is in development and will be launched in early 2015.

A searchable list of subject and domain specific data repositories can be found on the DataBib (<http://databib.org/>) and Registry of Research Data Repositories (<http://service.re3data.org>) websites.

If you intend to make data publicly available through a third party system, you should register its location in the LSHTM Research Data Repository, as outlined in the School Research Data Management Policy¹⁰. This will improve the likelihood that other researchers will be able to find your research in the long-term.

Example

During funding, digital data will be held securely in LSHTM's Teaming system, which will be accessible only to project members. LSHTM's SFX system will be used for secure file transfer, in cases where it is necessary to provide access to a 3rd party. Any research papers written and published during the funding period will be made available with a subset of the data necessary to verify the research findings, in compliance with the Wellcome Trust's OA Policy.

Following analysis and publication of findings, digital data (outputs 1, 2 and 4) will be deposited with the LSHTM Data Repository and made available to bona fide researchers for health-related research.

Further Information

- Wellcome Trust: Where Will You Make Data Available?
<<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#three>>
- LSHTM: Locate a Data Repository
<http://www.lshtm.ac.uk/research/researchdataman/share/locate_datarepository.html>
- UK Data Service: How to Share Data
<<http://ukdataservice.ac.uk/manage-data/plan/how-share.aspx>>

⁸ <http://researchonline.lshtm.ac.uk/612422>

⁹ <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/WTX060360.htm>

¹⁰ <http://researchonline.lshtm.ac.uk/612422/>

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

The Wellcome Trust encourages researchers to make their data available for access and use with as few restrictions as possible. However, they recognize that restrictions may need to be defined for certain types of data. Three questions should be considered:

- a. How will users learn of the existence of data and the content it contains?
- b. Who is allowed access to the data? Who mustn't/shouldn't be allowed to view the data and why?
- c. What measures should be applied to differentiate between users in these two groups?

First, you should consider how users will learn of the existence of your data. Few people will be interested in a set of unlabelled files published on a website. To attract interest, you should describe the content of your dataset. Many data repositories expect users to provide specific information about their data, such as its title, who created it, when and where it was captured, and so on.

Second, you should consider who is allowed access to the data. If you've identified specific groups that shouldn't be allowed to view the data (e.g. commercial companies) or need to limit how it is used (i.e. data can be used for non-commercial purposes or purposes compatible with the original research objectives only), control mechanisms will need to be established. These include:

- User accounts that limit read access to authorized users only;
- Procedures to monitor and evaluate access requests, e.g. user must complete a request form stating the variables that they require and the purpose for which they intend to use the data.
- Adoption of a Data Transfer Agreement that outlines conditions for access and use of the data

It may be appropriate to establish a graded access procedure, in which anonymised and aggregate data is made freely available, whereas sensitive and confidential data can only be accessed through application.

Example

To enable potential users to learn of the dataset's existence, structured metadata describing its content will be created and made available in human readable and machine processable form. The LSHTM Data Repository will publish metadata in several metadata formats, including Dublin Core, via OAI-PMH, RSS and ATOM, for indexing by search engines and harvesting by research data catalogues.

To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data.

Biological samples (output 3) will be deposited with the UK BioBank for future use, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points. Similar to the above, applicants will be required to comply with a Data Transfer Agreement prior to gaining access to data. DNA extracted from biological samples will be normalized and plated at Lab Y for use in future studies, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points.

Further Information

- Wellcome Trust: How Will Other Researchers be able to access Data?
<<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#four>>
- UK Data Service: Legal and Ethical <<http://ukdataservice.ac.uk/manage-data/legal-ethical.aspx>>
- LSHTM: Choose a Data Licence
<http://www.lshtm.ac.uk/research/researchdataman/share/choose_licence.html>

5. Do limits on data sharing need to be applied?

Outline any difficulties that will affect how data can be shared and used, along with the steps that will be taken to address them. Potential issues may include the need to:

- *Protect participant confidentiality:* Measures should be adopted to minimize the risk of identifying people participating in the study. If it is not possible to anonymise the data without losing its value, limits on who is permitted access may be defined.
- *Comply with informed consent agreement:* The Wellcome Trust expect consent procedures to include provision for data sharing, allowing it to be used for wider research. However, it recognizes that limitations upon permitted use may be necessary for health data. For example, participant consent may specify that data can only be used to investigate specific medical conditions or research questions.
- *Protect Intellectual Property Rights:* Health research often requires use of data that incorporates third party rights, e.g. government, funding bodies, research labs, and others. Steps should be taken to clarify rights associated with data and address any requirements.
- *Submit patent applications:* Projects that intend to apply for a patent as a result of their research may apply an embargo period to data sharing, only making it available after a specified time period.
- *Protect commercial confidentiality:* Research that involves a commercial partner or data that they have produced may require protection. Limitations may be established that restrict who is allowed access to the data and the permitted use.

Possible measures that may be applied to address these issues include: encryption of data during storage and transfer, anonymisation of personal information, development of Data Transfer Agreements that specify how data may be used by an end user, specification of embargo periods, and development of procedures and systems to limit access to authorized users only.

Example

To protect participant confidentiality, data outputs will be anonymised prior to deposit. To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule.

Biological samples are limited and depletable, so access will need to be carefully controlled and coordinated. The quantity of sample that is provided will be judged against the potential benefits of the research project, with advice from appropriate experts as required.

Further Information

- Wellcome Trust: Are any limits to data sharing required?
< <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#five> >
- UK Data Service: Legal and Ethical < <http://ukdataservice.ac.uk/manage-data/legal-ethical.aspx> >
- LSHTM: Data Sharing
< <http://www.lshtm.ac.uk/research/researchdataman/share/> >

6. What measures will be applied to preserve data?

Health data must be maintained for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule¹¹. To address these requirements, you should outline the activities that the project will perform to ensure data is accessible and usable in the long-term. These include:

- *Choose an appropriate file format:* The file format chosen should be supported by a wide range of software tools and capable of storing all required information at an appropriate level of granularity. The UK Data Service provide several recommendations:
<http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx>
- *Document your data:* Documentation is essential to ensure data can be understood and used by yourself, your colleagues and other users. Health researchers are often required to apply domain-specific standards to describe their digital resources. For example:
 - DDI metadata standard (<http://www.ddialliance.org/>) for social science data
 - SDMX standard (<http://sdmx.org/>) for statistical data
 - Clinical Standard Operating Procedures (<https://intra.lshtm.ac.uk/trials/sops/>)

Advice on documentation & metadata standards is available through the LSHTM RDM Support Service¹².

Example

Data will be provided in file formats considered appropriate for long-term access, as recommended by the UK Data Service (<http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx>). For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule.

Appropriate documentation necessary to understand the data will also be provided. This will include high-level information on the study itself and a comprehensive data dictionary, which describes the purpose of each variable and the permitted values.

Further Information

- Wellcome Trust: How will you ensure that key datasets are preserved to ensure their long-term value?
< <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#six> >
- UK Data Service: Document your Data
< <http://ukdataservice.ac.uk/manage-data/document.aspx> >
- LSHTM: Documenting your Data
< <http://www.lshtm.ac.uk/research/researchdataman/describe> >

¹¹ <https://intra.lshtm.ac.uk/informationmanagement/records/retention.html>

¹² http://www.lshtm.ac.uk/research/researchdataman/rdm_service.html

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

Finally, you should outline any resources that you will need to manage and share your research data. Traditionally, researchers have been wary of stating these costs, believing that it will make their application appear uncompetitive. However, many funders view the lack of data management costs as an area for concern, potentially suggesting that the applicant has not considered the practicalities of their project.

In an RCUK briefing on data management and sharing costs¹³, the Wellcome Trust offered the following advice on resource allocation:

1. *Resources should be clearly described and justified in research applications:* Sufficient information should be provided to help the evaluator to identify resources allocated for data management, the reason that it is needed and where costs arise.
2. *Institutional services must not be claimed for twice:* Institutional resources that are already claimed in indirect costs must not also be included in direct costs.
3. *Post-project costs must be spent before project end:* Costs for long-term curation and preservation, such as POSF (Pay Once, Store Forever) storage, must be allocated before the close of the project grant. The WT will not support unexpected costs incurred after the project end.

Specific resources that the Wellcome Trust encourages applicants to consider include:

- *People and Skills:* Do project staff have the necessary time and expertise to prepare data for sharing? Are additional specialist staff required, e.g. to create metadata for the dataset?
- *Training:* Have training needs been identified and costed?
- *Technical Infrastructure:* Does the project have access to systems capable of storing, managing and sharing data during the project lifetime and following its completion? An overview of LSHTM storage systems can be found at <http://www.lshtm.ac.uk/research/researchdataman/store/> and <http://www.lshtm.ac.uk/its/staffservices/>.

Example

To implement the stated plan, the project team will consult LSHTM's IT Services for advice on data storage and the RDM Support Service for training on metadata creation. In the final six months, the Data Manager will prepare data for sharing. Activities will include anonymisation of personal information, export to an open format and creation of descriptive metadata. They will meet with LSHTM's RDM Support Service at appropriate intervals during this time to confirm procedures for deposit to the LSHTM Data Repository and verify that work performed is fit for purpose.

Further Information

- Wellcome Trust: What resources will you require to deliver your plan?
<<http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#seven>>
- RCUK: Supporting research data management costs through grant funding
<<http://blogs.rcuk.ac.uk/2013/07/09/supporting-research-data-management-costs-through-grant-funding/>>
- UK Data Service: Plan to Share
<<http://ukdataservice.ac.uk/manage-data/plan.aspx>>

¹³ <http://blogs.rcuk.ac.uk/2013/07/09/supporting-research-data-management-costs-through-grant-funding/>

Sample Data Management & Sharing Plan

By answering each question raised in the Wellcome Trust guidance, you will begin to build a Data Management & Sharing Plan suitable for submission. You're likely to have repeated the same information for multiple questions, so some amount of de-duplication and re-organisation will be necessary.

The worked example below illustrates how these various components can be brought together to form a coherent Data Management & Sharing Plan.

Worked Example

Data capture will be performed during months 8-20, at which point four types of data will be collected from the (approximately) 500 study subjects:

1. *Questionnaire*: An interviewer-administered CAPI questionnaire will be performed covering medical history, health and socio-demographic circumstances
2. *Clinical*: A clinical examination will be conducted to measure blood pressure, anthropometry, 12-lead digital ECG assessed for abnormalities; assessment of physical functions (grip strength, chair rise, walk speed) and cognition.
3. *Biological samples*: Blood samples will be taken from each subject and centrifuged within 2 hours. Together with serum and plasma, it will aliquoted into 20 bar-coded cryovials and stored at -70C
4. *Interviews*: It has previously been found that participant's partners tend to report higher levels of alcohol consumption and more frequent episodes of hazardous drinking than men themselves. We will interview partners of 1 in 4 of the study participants (selected randomly) in order to estimate the extent of under or over reporting of behaviours.

All 4 data types will be usable for future research in some form, subject to appropriate measures being implemented to protect participant confidentiality. Digital and physical outputs will be made available to bona fide researchers for health-related research, irrespective of their institution (university, charity, government, commercial) or location (UK or elsewhere).

During funding, digital data will be held securely in LSHTM's Teaming system, which will be accessible only to project members. LSHTM's SFX system will be used for secure file transfer, in cases where it is necessary to provide access to a 3rd party.

Digital data (outputs 1, 2 and 4) will be deposited with the LSHTM Data Repository and made available to bona fide researchers for health-related research. The study team will make digital data (outputs 1, 2 and 4) available within 6 months of project completion. This embargo period is requested to allow time for additional analysis and further publication of research findings to be performed. Research papers written and published during the funding period will be made available with a subset of the (anonymised) data necessary to verify the research findings, in compliance with the Wellcome Trust's OA Policy.

To protect participant confidentiality, data outputs will be anonymised prior to deposit. To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule, although there is no reason that it cannot be held indefinitely.

Data will be provided in file formats considered appropriate for long-term access, as recommended by the UK Data Service (<http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx>). For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview

transcripts. Appropriate documentation necessary to understand the data will also be provided. This will include high-level information on the study itself and a comprehensive data dictionary, which describes the purpose of each variable and the permitted values.

To enable potential users to learn of the dataset's existence, structured metadata describing its content will be created and made available in human readable and machine processable form. The LSHTM Data Repository will publish metadata in several metadata formats, including Dublin Core, via OAI-PMH, RSS and ATOM, for indexing by search engines and harvesting by research data catalogues.

To gain access, researchers will be required to complete a data request form, stating the purpose for which they intend to use it. If this complies with the research objectives of the original research, they will be asked to sign a Data Transfer Agreement stating that they will not make any attempt to identify participants, among other requirements. If they agree to these conditions, they will be provided with a copy of the requested data.

Biological samples (output 3) will be deposited with the UK BioBank for future use, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points. As these samples are limited and depletable, access will need to be carefully controlled and coordinated. The quantity of sample that is provided will be judged against the potential benefits of the research project, with advice from appropriate experts as required. Similar to the above, applicants will be required to comply with a Data Transfer Agreement prior to gaining access to data. DNA extracted from biological samples will be normalized and plated at Lab Y for use in future studies, e.g. to assess the effects of alcohol on biomarkers and risk of cardiac damage to surrogate end-points.

To implement the stated plan, the project team will consult LSHTM's IT Services for advice on data storage and the RDM Support Service for training on metadata creation. In the final six months, the Data Manager will prepare data for sharing. Activities will include anonymisation of personal information, export to an open format and creation of descriptive metadata. They will meet with LSHTM's RDM Support Service at appropriate intervals during this time to confirm procedures for deposit to the LSHTM Data Repository and verify that work performed is fit for purpose.

Further Guidance

Guidance on Wellcome Trust requirements can be found at:

- Wellcome Trust: Policy on data management and sharing
< <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm> >
- Wellcome Trust: Developing a Data Management and Sharing Plan
< <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm> >
- London School of Hygiene & Tropical Medicine: Wellcome Trust guidance
< http://www.lshtm.ac.uk/research/researchdataman/plan/funder_wellcome.html >
- University of Bristol: Data Management Planning guide for Wellcome Trust applicants
< <http://data.bris.ac.uk/research/dmp/dmp-nonrcuk/> >
- University of Exeter: A Guide to Wellcome Trust Data Management and Sharing Plans
< <http://as.exeter.ac.uk/library/resources/rdm/create/datamanagementplans/funderguidance/> >