Guidance Notes for Completing a "Checklist for a Data Management Plan v3.0"

For Researchers in the Psychological Sciences

Dr Richard R. Plant
r.plant@psych.york.ac.uk

DMTpsych Project
Department of Psychology, The University of York,
York, YO10 5DD, UK

Funded by JISC

Background

Funding bodies increasingly require grant-holders to develop and maintain Data Management Plans (DMPs) to keep track of the stewardship of their often irreplaceable digital resources. Such plans cover maintaining research data together with any other digital resources a project produces. The ultimate goal is to produce a plan which covers not only maintenance but also allows for data and resource sharing within the wider academic community. Such plans are now a mandatory requirement for obtaining funding from many research councils and often must be submitted with the research grant application itself.

To help researchers across all disciplines the JISC funded Digital Curation Centre based in Edinburgh (UK) has produced a number of tools to enable creation of exemplar Data Management Plans. Outputs from these tools can be used to meet the requirements of funders.

> *"Having analysed UK research funders' requirements for Data Management and Sharing Plans, (see details under Policy and Legal) the DCC has created the Checklist for a Data Management Plan, a comprehensive list of the details that researchers may be asked to include in such plans."*
>
> *"A flexible web-based tool has also been developed, based on the latest version of the Checklist. DMP Online assists users in the planning process by enabling them to create personalised plans according to their context or research funder."*
>
> *The DCC (http://www.dcc.ac.uk/resources/data-management-plans)*

This guidance booklet is intended to provide discipline specific orientation and practical examples on how to complete the "DCC's Checklist for a Data Management Plan" if you are a researcher in the Psychological Sciences. Ultimately by completing either the paper based or online version of the checklist you will end up with a Data Management Plan that meets most funders requirements.

Each question on the "Checklist for a Data Management Plan" is mirrored by an entry in this guide. Each entry provides discipline specific orientation followed by a practical example based on a real world scenario.

Practical examples given in this guide are taken from the standpoint of a researcher attempting to complete a Checklist. Examples in the main are structured as though J. Ridley Stroop were applying to a modern funder to obtain a grant to fund his 1935 work on the now famous Stroop effect (Journal of Experimental Psychology, 18, 643-662 [http://psychclassics.yorku.ca/Stroop/]). In this scenario he is also applying as though from a modern department.

Practical examples of responses to questions in the Checklist are given in the grey shaded areas. You should bear in mind that examples in this guide may be fuller than those you may need to make in order to complete a checklist to accompany your own research grant application. However you should try to give as much detail as possible.

Dr Richard R. Plant, July 2011

If you would like to take part in a Stroop experiment yourself visit: http://cognitivefun.net/test/2

1. Introduction and Context

1.1 Basic project information

1.1.1. Name of project

The name of the project together with any acronym used. If the project has been funded you should also include the funding reference. If the project is at the application stage you should include a reference number or the funding strand of the call where available.

Interference in serial verbal reactions - ISVR (ESRC/10012/2011/AS)

1.1.2. Funding body (or bodies)

List the funding body, or bodies, in order of funding amount with the major funder listed first. Include the official funding body name and address.

Economic and Social Research Council (ESRC)
Polaris House
North Star Avenue
Swindon
SN2 1UJ

1.1.3. Budget

The total budget to be provided over the length of the grant.

£60,000

1.1.4. Duration

Enter the number of months or years the project will run from the proposed start date.

12 months

1.1.5. Lead partner organisation

The lead partners name and official contact address.

Dr J Ridley Stroop
Department of Psychology
The University of York
York
YO10 5DD
UK

1.1.6. Other partner organisations

List any other partner organisations in order of their contribution in resources to the project, either in terms of time or financial contribution, with those contributing most listed first. Include collaborators names and official contact addresses if known.

None

1.2. Short description of the project's fundamental aims and purpose

Most grant application forms have a cover sheet, or short project summary section, which succinctly summarises the project's outline and goals. Generally this should be included verbatim.

Please note that the abstract below has been substantially re-worded in order to "recreate" an example grant application based on the subsequently published work of J. Ridley Stroop (1935).

The difference in time for naming colors and reading color names has been variously explained. Cattell (1886) and Lund (1927) have attributed the difference to 'practice'. Woodworth and Wells (1911, p. 52) have suggested that, "The real mechanism here may very well be the mutual interference of the five names, all of which, from immediately preceding use, are 'on the tip of the tongue,' all are equally ready and likely to get in one another's way." Brown (1915, p. 51) concluded "that the difference in speed between color naming and word reading does not depend upon practice" but that (p. 34) "the association process in naming simple objects like colors is radically different from the association process in reading printed words."

The present problem grew out of experimental work in color naming and word reading conducted in Jesup Psychological Laboratory at George Peabody College For Teachers. The time for reading names of colors had been compared with the time for naming colors themselves. This suggested a comparison of the interfering effect of color stimuli upon reading names of colors (the two types of stimuli being presented simultaneously) with the interfering effect of word stimuli upon naming colors themselves. In other words, if the word 'red' is printed in blue ink how will the interference of the ink-color 'blue' upon reading the printed word 'red' compare with the interference of the [p. 647] printed word 'red' upon calling the name of the ink-color 'blue?' The increase in time for reacting to words caused by the presence of conflicting color stimuli is taken as the measure of the interference of color stimuli upon reading words. The increase in the time for reacting to colors caused by the presence of conflicting word stimuli is taken as the measure of the interference of word stimuli upon naming colors. A second problem grew out of the results of the first. The problem was, What effect would practice in reacting to the color stimuli in the presence of conflicting word stimuli have upon the reaction times in the two situations described in the first problem?

The proposed project will investigate five areas within interference in serial verbal reactions:

1. Interference in serial verbal reactions will be studied by means of newly devised experimental materials. This will test whether the source of the interference is in the materials themselves. The words red, blue, green, brown, and purple will be used on a test sheet. No word will be printed in the color it names but in an equal number of each of the other four colors; i.e. the word 'red' is printed in blue, green, brown, and purple inks; the word 'blue' is printed in red, green, brown, and purple inks; etc. Each word presents the name of one color printed in ink of another color. Hence, a word stimulus and a color stimulus both are presented simultaneously. The words of the test will be duplicated in black print and the colors of the test will be duplicated in squares or swastikas. The difference in the time for reading the words printed in colors and the same words printed in black is the measure of the interference of color stimuli upon reading words. The difference in the time for naming the colors in which the words are printed and the same colors printed in squares (or swastikas) is the measure of the interference of conflicting word stimuli upon naming colors.

2. The interference of conflicting color stimuli upon the time for reading 100 words (each word naming a color unlike the ink-color of its print) is predicted to cause in increase in response time over the normal time for reading the same words printed in black. These tests provide a unique basis (the interference value) for comparing the effectiveness of the two types of associations.

3. As a test of the permanency of the interference of conflicting word stimuli to naming colors eight days of practice (200 reactions per day) will be given in naming the colors of the print of words (each word naming a color unlike the ink-color of its print). The effects of this practice will then be assessed in terms of: (a) Decrease in interference of conflicting word stimuli to naming colors; (b) Production of a practice curve similar to that observed in other learning experiments; (c) An increase invariability; (d) Shorter reaction time to colors presented in color squares and;

(e) Increased interference of conflicting color stimuli upon reading words.

4. Whether practice effects the variability of the group depending upon the nature of the material used.

5. If sex differences in naming colors is due to the difference in the training of the two sexes.

1.3. Related policies

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

This will vary according to the funder, or funders, but will typically include information on both "data" management and "data" sharing. Where practicable you should include a succinct summary, or if this isn't appropriate, you should acknowledge the funders guidance and provide a reference to their policy, e.g. a weblink or include a full version in an annex to this plan. The required information can often be found on the funders website or may be contained within the funding call documentation. If you have more than one funder you should include summary notes and appropriate references.

> *Definition of "data": In the terms of the Data Protection Act 1998, data are information relating to an individual where the structure of the data allows information about the individual to be readily accessed. The information may be held in manual form (e.g., as written notes relating to a person or as part of a filing system, including card index or filing cabinets structured by name, address or other identifier) or in a form capable of being processed electronically. Personal data are any data relating to a living individual (e.g., name, address, payroll details, exam results). Sensitive data form a subset of personal data that relate to a living person, recording such things as racial or ethnic origin, political opinions, religious beliefs, trade union membership, health, criminal convictions, etc. Data are processed whenever compiled, stored or otherwise operated upon. So disseminating the examination results of students involves processing data relating to each of them, as does giving and receiving personal references, producing agenda items or minutes for committees at which students are discussed as individuals, etc. Similarly, data about staff are processed when they are committed to manual or electronic records held within the institution.*

We will adhere to the "ESRC Research Data Policy" guidelines relating to the responsibilities of grant applicants (September 2010) as laid out below. A full copy can be found in Annex X or at http://www.esrc.ac.uk/ESRCInfoCentre/Images/ESRC_Research_Data_Policy_2010_tcm6-37350.pdf

Responsibilities of ESRC grant applicants

Those ESRC grant applicants who plan to generate data are responsible for preparing and submitting data management and sharing plans for their research projects as an integral part of the application.

It is expected that an outline data management and sharing plan will include the following points:

- an explanation of the existing data sources that will be used by the research project with references;
- an analysis of the gaps identified between the currently available and required data for the research;
- information on the data that will be produced by the research project, including the following:
    - data volume
    - data type, e.g. qualitative or quantitative data
    - data quality, formats, standards documentation and metadata
    - methodologies for data collection

- planned quality assurance and back-up procedures [security/storage];
- plans for management and archiving of collected data;
- expected difficulties in data sharing, along with and causes and possible measures to overcome these difficulties;
- explicit mention of consent, confidentiality, anonymisation and other ethical considerations;
- copyright and intellectual property ownership of the data; and
- responsibilities for data management and curation within research teams at all participating institutions.

The ESRC requires that all applicants seeking ESRC funding include a statement on data sharing in the relevant section of the Je-S application form. If data sharing is not possible, the applicant must present a strong argument to justify their case. The ESRC reserves the right to decline the request or demand additional information from the applicant.

Responsibilities of ESRC grant holders

It is a responsibility of the award holder to incorporate data management and sharing as an indivisible part of the research project to increase the potential for data to be shared. This should address potential issues of confidentiality, ethical issues, legal issues, time constraints and other issues which could limit data sharing opportunities from the very start of the project.

All ESRC funded research projects, collecting or producing data, are required to develop and implement a data management plan to ensure that data are well managed during their life-cycle and are ready to be offered for archiving and sharing when a project ends. Planning for data sharing should be done at the earliest stages of project design and well in advance of beginning fieldwork, particularly if it includes the collection and subsequent management of potentially confidential data that may affect data sharing. The award holder shall seek advice and guidance from the ESRC Economic and Social Data Service (ESDS) at the outset of the project to clarify how issues of confidentiality and sharing are to be addressed.

The award holder is expected to report on the on-going implementation of the data management and sharing plan through annual reporting to ESRC. In cases where there are ad hoc issues that may potentially have an impact on data sharing, the award holder must immediately raise these with the respective ESRC Case Officer so that further guidance and support can be provided.

It is the responsibility of the award holder to formally offer any data created or repurposed during the lifetime of the award to the ESDS within three months of the end of the award. The award holder is responsible for providing these data to the ESDS for assessment, and if accepted, to ensure that they meet the requirements of the ESDS for preservation and future re-use. If data were accepted, the award holder is expected to make them available to the ESDS for preparation for re-use and archiving without delay.

Sharing sensitive data – joint responsibilities

Most data generated as a result of economic and social research can be successfully archived and shared. The ESRC recognises that some research data are more sensitive than others and argues that it is a responsibility of the award holders to consider all issues related to confidentiality, security and copyright before initiating the research.

Where research data are considered confidential or contain sensitive personal data, award holders must seek to secure consent for data sharing or alternatively anonymise the data in order to make sharing possible. The ESRC regards a waiver of deposit as an exception and reserves the right to refuse waivers where there is insufficient evidence to prevent archiving and data sharing.

Where issues of confidentiality are foreseen that would prevent data being successfully shared, the award holder is encouraged to contact the ESDS at the earliest opportunity. The ESDS will provide all necessary support to the award holders, guiding them through various strategies to enable data sharing. The ESRC supports the position that sensitive and confidential data can be shared ethically provided researchers pay attention right from the planning stages of research to the following aspects:

- when gaining informed consent, include consent for data sharing;
- where needed, protect participants' identities by anonymising data; and
- address access restrictions to data before commencing research in the data management and sharing plan

Only where researchers have demonstrated due diligence in all three areas will waivers be granted (http://www.esds.ac.uk/aandp/create/dataman.asp).

Intellectual Property Rights
The ESRC endorses the Research Councils position on the exploitation of research results and therefore positively encourages the exploitation of the results of research the ESRC support, as a contribution to enhance the quality of life, sustainability and competitiveness of the UK.

In respect of research grant funding, unless stated otherwise, the ownership of intellectual property and responsibility for its exploitation, rests with the organisation carrying out the research. The ESRC may, in specific cases, reserve the right to retain ownership of the intellectual property and to arrange for it to be exploited for the national benefit in other ways. If exercised, this condition is included in the terms of the relevant award. On the exploitation of research results and therefore positively encourages the exploitation of the results of research the ESRC support, as a contribution to enhance the quality of life, sustainability and competitiveness of the UK.

In taking responsibility for exploiting intellectual property, the ESRC expects the research organisation to ensure that individuals associated with the research understand the arrangements for exploitation. Where research is funded by or undertaken in collaboration with others, the research organisation is responsible for putting appropriate formal agreements in place covering the contributions and rights of the various organisations and individuals involved. Such agreements must be in place before the research begins. Research organisations are required to ensure that the terms of collaboration agreements do not conflict with the Terms and Conditions for Research Council Grants.

Copyright and Confidentiality
The ESRC expects award holders to meet the copyright requirements set down in the Copyright, Designs and Patents Act 1998. Responsibility for ensuring compliance with all laws and other legal instruments rests with the award holders and/or their institutions. The ESRC will not accept liability for any complaint or legal action taken against a researcher or the ESDS for infringements of copyrights, defamation or any other data protection requirements.

Security
The ESRC promotes and adheres to relevant levels of information security. The Council is committed to make all possible arrangements to ensure the protection of data from unauthorised use, change, disclosure or destruction.

The ESRC Secure Data Service has been established to promote excellence in research by enabling safe and secure remote academic access to data hitherto deemed too sensitive, detailed, confidential or potentially disclosive to be made available under standard licensing and dissemination arrangements. Data which can be anonymised without loss of significant informational content are not appropriate for the Secure Data Service.

Data Protection and Freedom of Information
The ESRC expects award holders to adhere to the Data Protection Act 1998, which contains eight (enforceable) principles of good practice, applying to anyone processing personal data, including the use of personal data in research. These include obtaining the data subject's consent or meeting at least one of the 'necessary' conditions described in the Act.
The ESRC complies with the requirements of the Freedom of Information Act 2000 that establishes a general right of access to all types of recorded information held by public authorities, including Government Departments and Non-Departmental Public Bodies (http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/about/CI/freedom_of_information/FoI_Intro

duction.aspx).

If the Principal Investigator does not state to the contrary in the Je-S application form, it will be assumed that they are willing for their contact details and other relevant information to be shared with the relevant data service provider working with the ESRC (http://www.esds.ac.uk/aandp/create/dataman.asp).

Ethical considerations
The ESRC, in facilitating innovative and high quality research, requires that the research it supports will be carried out to a high ethical standard. ESRC award holders are, therefore, required to adhere to the key principles of ethical research addressed in the ESRC Framework for Research Ethics (http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/opportunities/research_ethics_framework/).

1.3.2. Institutional or research group guidelines

This will vary according to the institution, or department you are working in but will typically include information on both "data" management and "data" sharing. Where practicable you should include a succinct summary or if this isn't appropriate you should acknowledge the institution or department guidance and provide a reference to their policy, e.g. a weblink or include a full version in an annex to this plan. The required information should be available from the institutions Research Office or Experimental Officer within your department or on its website. If these responsibilities are devolved to individual research groups then they should be approached for guidance and policy statements.

Department of Psychology, University of York: Data Storage and Archiving Policy

Introduction
The research councils require research data to be kept in a secure manner while in active use and also for the raw data be archived for a period of between five and twenty years on the completion of a project or submission of a paper. It is the responsibility of the institution at which the research was carried out to retain the archive data. This policy outlines the steps that must be taken to comply with these requirements and provides explanatory notes.

Policy

- Active data must be stored in a secure fashion.
- Active data must be backed up on a regular and frequent basis.
- A backup copy of all data must be stored offsite on at least a weekly basis.
- The Department will make available a suitable storage system and will be responsible for the backup of data stored on said system.
- On submission of a paper, the raw data must be submitted to the Department for archiving.
- The Department will securely store archived data for the required period of time and make it available as appropriate.
- All Departmental/University administrative data must be stored on the fileserver.

Explanatory Notes
1. Data must be stored on a system that:

- is kept up to date with security patches and updates. All systems must be set to obtain anti-virus and operating system updates automatically. Systems for which this is not appropriate must either have a formal schedule defined for manual updates to be applied or not be connected to the network. Note that Microsoft do not provide updated for systems older than Windows 2000.
- requires a password to access. Data must not be stored on open shares (ones that can be accessed without a password), or on computers which do not require a password to access. The use of shared passwords is strongly discouraged due to the lack of accountability and should be avoided wherever possible.
- has redundancy built in to guard against data loss. Such redundancy must be designed to

prevent against the failure of a single system (or component of a system) causing the loss of data.

It should be noted that data kept on physical media must also be kept in a secure fashion.

2. Any system holding research data must have an associated backup facility. This must back up onto either removable media or to a different system.

3. All data must have a backup copy stored off site at least once a week. This means that using a single external hard drive is insufficient unless the data is transferred onto another system off site. If a copy of data is kept on a live system offsite, this system should also meet the security and redundancy requirements set out in point 1 and must meet the password requirement.

4. With the purchase of the central storage array from Phase 1 of the IT Strategy, the Department can now offer secure, reliable storage with regular and offsite backup.

- The Department will provide 20GB storage per user.
- The Department will also offer an additional 20GB storage per research group. This can be used to share files between researchers working on a project or to transfer data from experimental systems to an individual researcher's filespace.
- Additional storage can be purchased at the cost of £10/GB for a 3 year period or £5/GB for a single year. This charge would cover:
- Provision of the purchased amount of storage.
- Setting security permissions to allow access either to an individual/lab group or named persons.
- Provision of instructions for mounting the storage from individual PCs or Macs.
- Daily/weekly/monthly backups.
- Recovery of deleted files (either from backup or shadow copies)
- This cost should be considered against the cost of:
- Purchase of additional secure storage.
- Ongoing time requirements of synchronisation of files between local systems and this storage.
- (At least) Weekly transfer of this data off site.
- Ongoing maintenance of security on local systems.

5. On submission of a paper, all relevant data must be submitted to the Department. This should include:

- Information about the paper (title, journal, authors…).
- The raw data.
- An index to the data files if required (such as a mapping between data files and participants).
- A description of the structure of the data files.
- If required, a key to the meaning of triggers of experimental conditions within the data file.
- Stimuli.
- Optionally, any additional information (such as processed data, figures or a copy of the paper).

6. All of the information detailed in point 5 will be archived into an individual Zip file and stored on two copies of archive media along with checksum information. These will be stored in different locations. On an annual basis, the media will be checked for consistency.

The Department will trial a system over the coming months to ensure that all relevant data is stored. An online database will be constructed to allow external users to find and request the data. When this is complete, the URL will be disseminated to allow researchers to publish it along with papers.

7. All Departmental administrative files (support staff files, tutorial data, examination

papers/results…) must be stored on the Departmental file server in an area with security permissions appropriate to the data.

1.3.3. Other policy-related dependencies

In addition to funder and institutional, or departmental, requirements there may be governmental acts you need to comply with. Typically as a minimum you will need to comply with the Data Protection Act 1998 and Freedom of Information Act 2000. If you have not covered compliance to both acts in the funders and institutional/departmental sections above you can include them here. If you work in a clinical setting, or within an NHS trust, you may also need to reference Caldicott principles (http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_114509). For further information you are advised to consult the "Caldicott Guardian" either at your local institution or the trust concerned.

If your project stores data about an "identifiable" individual you are likely to be processing personal data and should comply with the Data Protection Act 1998, e.g. experimental results, test scores, clinical records etc. which can be linked to a living person. This applies both to data held on computers and to data held manually, i.e. on paper. If your data is anonymised you should say that you do not need to fully comply with the Data Protection Act for this reason but intend to follow it's principles. For more information on the DPA 1998 consult the Information Commissioner's Office (http://www.ico.gov.uk/). In addition you should comply with the requirements of the Freedom of Information Act 2000 which provides, amongst other things, guidance on how long data should be stored for and how to respond to requests for information (http://www.ico.gov.uk/). If your project needed to obtain the approval of an ethics committee then it is likely that you will have addressed issues of data confidentiality. Where ethical approval has been sought and granted you should include a summary of the requirements you need to adhere to or a reference to them. For background on the Freedom of Information Act 2000 see: http://www.ico.gov.uk/Global/faqs/freedom_of_information_for_organisations.aspx

It is likely that the institution in which you are based is registered with the Information Commissioner's Office and that there is a named person, or department, who deals with both the Data Protection and Freedom of Information Acts and will be able to advise you where needed. Their contact information should be listed on your institutions website or contact your local Research Support Office.

Data Protection
The Studies of interference in serial verbal reactions project will adhere to the eight data protection principles that govern the use of personal information as described in the Data Protection Act 1998. Any personal information held by the project will be:

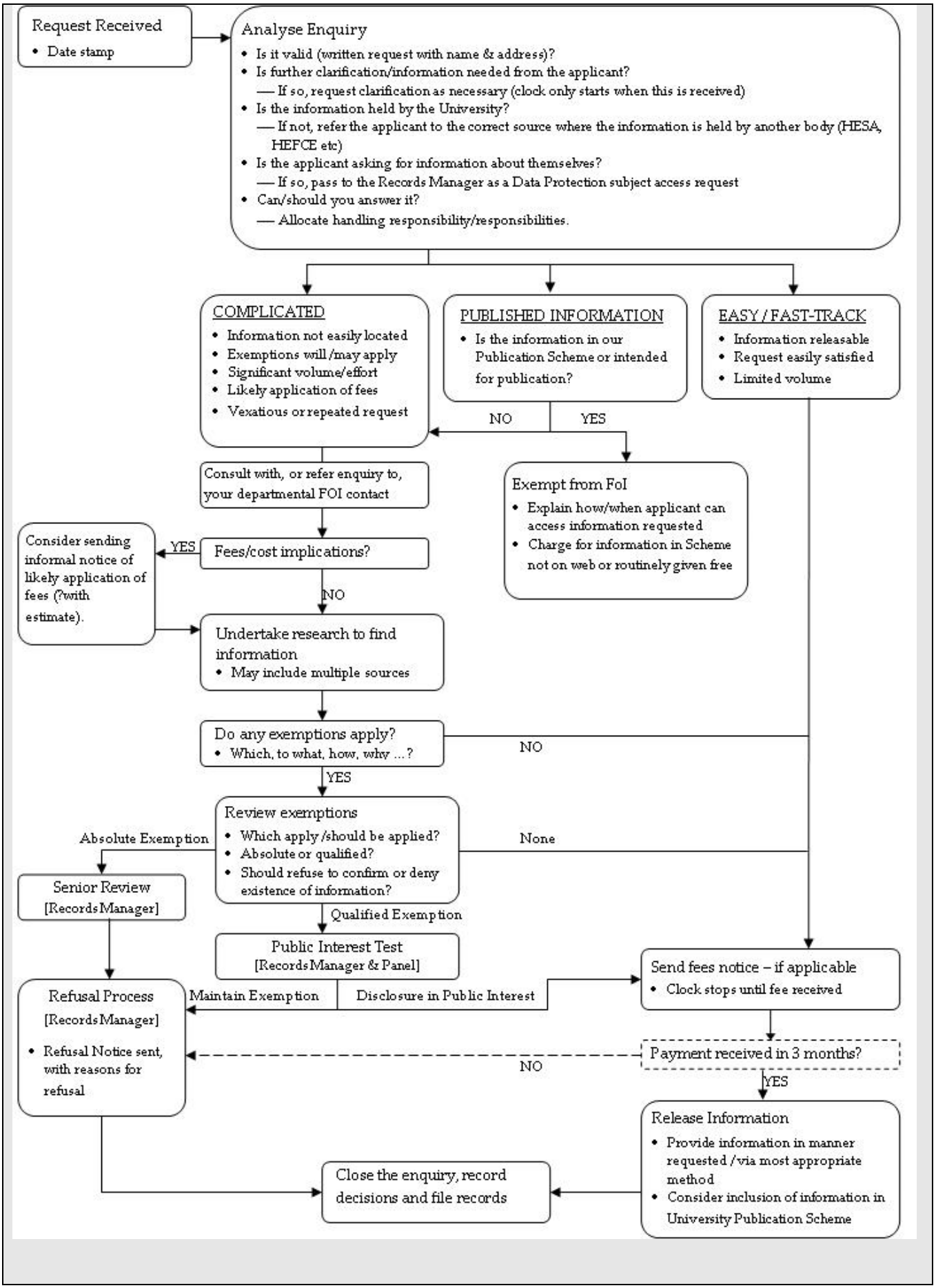
- processed fairly and lawfully
- processed for one or more specified and lawful purposes, and not further processed in any way that is incompatible with the original purpose
- adequate, relevant and not excessive
- accurate and, where necessary, kept up to date
- kept for no longer than is necessary for the purpose for which it is being used
- processed in line with the rights of individuals
- kept secure with appropriate technical and organisational measures taken to protect the information
- not transferred outside the European Economic Area (the European Union member states plus Norway, Iceland and Liechtenstein) unless there is adequate protection for the personal information being transferred

The institutions registration and notification of its uses of personal data can be viewed as part of the Public Register of Data Controllers maintained by the Information Commissioner's Office (http://www.ico.gov.uk/what_we_cover/promoting_data_privacy/keeping_the_register.aspx). The University's registration number is Z4855807. Contact the Records Management office for further information (http://www.york.ac.uk/recordsmanagement/).

Freedom of Information

The Studies of interference in serial verbal reactions project will comply with the Freedom of Information Act 2000 and will handle any requests within a timely manner as described in the following document and represented graphically below:

http://www.york.ac.uk/recordsmanagement/foi/enquiryguide.htm


Request Received
• Date stamp
Analyse Enquiry
• Is it valid (written request with name & address)?
• Is further clarification/information needed from the applicant?
— If so, request clarification as necessary (clock only starts when this is received)
• Is the information held by the University?
— If not, refer the applicant to the correct source where the information is held by another body (HESA, HEFCE etc)
• Is the applicant asking for information about themselves?
— If so, pass to the Records Manager as a Data Protection subject access request
• Can/should you answer it?
— Allocate handling responsibility/responsibilities.
COMPLICATED
• Information not easily located
• Exemptions will/may apply
• Significant volume/effort
• Likely application of fees
• Vexatious or repeated request
PUBLISHED INFORMATION
• Is the information in our Publication Scheme or intended for publication?
EASY / FAST-TRACK
• Information releasable
• Request easily satisfied
• Limited volume
NO
YES
Consult with, or refer enquiry to, your departmental FOI contact
Exempt from FoI
• Explain how/when applicant can access information requested
• Charge for information in Scheme not on web or routinely given free
Consider sending informal notice of likely application of fees (?with estimate).
YES
Fees/cost implications?
NO
Undertake research to find information
• May include multiple sources
Do any exemptions apply?
• Which, to what, how, why ...?
NO
YES
Review exemptions
• Which apply/should be applied?
• Absolute or qualified?
• Should refuse to confirm or deny existence of information?
Absolute Exemption
None
Senior Review
[Records Manager]
Qualified Exemption
Public Interest Test
[Records Manager & Panel]
Refusal Process
[Records Manager]
• Refusal Notice sent, with reasons for refusal
Maintain Exemption
Disclosure in Public Interest
Send fees notice – if applicable
• Clock stops until fee received
Payment received in 3 months?
NO
YES
Release Information
• Provide information in manner requested/via most appropriate method
• Consider inclusion of information in University Publication Scheme
Close the enquiry, record decisions and file records

1.4. Basic Data Management Plan information

1.4.1. Date of creation

Enter the date the plan was completed.

01/10/2010

1.4.2. Aims and purpose of this plan

The main aims of any data management plan should be broadly similar. These typically encompass data collection, standards, quality controls, storage, access and dissemination and sharing along with Intellectual Property considerations. It is possible that the institution or NHS trust you are working in may provide standardised answers for this question.

The purpose of this plan is to is to specify the following in relation to the Studies of interference in serial verbal reactions project:

- Data collection policies
- Data standards
- Data quality and quality assurance
- Short and long term data storage
- Policies on access to data
- Intellectual Property considerations
- Data dissemination and sharing

1.4.3. Target audience for this plan

The target audience for the Data Management Plan is generally the funder(s), the project manager/principal investigator and/or person responsible for data management and the institution(s). In specific applications, such as clinical work, this may include the Caldicott Guardian who is based within the NHS trust where the work will take place. For example funders such as ESRC may have their own Data Archive (http://www.data-archive.ac.uk/) where you may be required to deposit data as a condition of funding. If named repositories exist these should be listed along with the funders details. It is possible that the institution or NHS trust you are working in may provide standardised answers for this question.

The target audience for this Data Management Plan comprises:

- The funding body ESRC
- The project manager/principle investigator and/or person responsible for data management
- The Records Management Office at the University of York who are the named lead institution (http://www.york.ac.uk/recordsmanagement/)
- Any other interested parties

2. Data Types, Formats, Standards and Capture Methods

2.1. Give a short description of the data being generated or reused in this research

Ideally in this section you should give a short description of the data you will generate and/or reuse. It is accepted that over the course of the project these estimates may change. Where possible you should give an indication of the size or volume of the final dataset, e.g. in GigaBytes for computer data, completed paper questionnaires, or Dictaphone tapes. At this stage it is acceptable to use estimates as a plan should be revisited, revised and refined as the project progresses.

> The data generated by this scientific experiment will mainly be numerical reaction times recorded over a number of trials and sessions per participant. Numerical error rates will be recorded. Biodata such as gender, age etc. will also be recorded. It is anticipated that raw data will be in the order of 20Mb per participant over the length of the project and that there will be approximately 100 participants. In total it is estimated that this will generate a final dataset of around 2Gb.

2.2. Existing Data

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

A yes/no response is required. In most circumstances you will answer "yes" as checking whether new data need to be generated will be a cornerstone of your funding application.

> Yes.

2.2.2. If Yes, What existing datasets could you use or build upon? (Contingent on 2.2.1 above)

If existing datasets exist that can be reanalysed you are encouraged to mention them here and briefly how you intend to use them. In the majority of cases your study will be a variation on an existing study or completely novel. It is also likely that you will not have access to raw data produced by other researchers' even if an identical study to the proposed one exists. It is possible however that reanalysis of clinical data, neuroimaging data or other meta-analysis means that your study may not generate any truly new data. In such cases you will still be generating "new" data as a result of new analyses and it is this data will need to be included in the Data Management Plan.

> We have examined the datasets produced by authors such as Lund (1927) and Ligon (1932) which tested a similar hypothesis to our own but with a younger age range of subjects.

2.2.3. Are there any access issues pertaining to the pertinent, existing data? (YES/NO Contingent on 2.2.1 and 2.2.2 above - Include financial costs if relevant)

A yes/no response is required. In this section you are free to mention aspects that prevent you accessing any pertinent existing data. You can also mention costs in this section. For example you could say that you could gain access only to published data and not to raw data from a given study.

> No.

2.3. New Data

2.3.1. Why do you need to capture/create new data?

Generally most research projects will capture or generate new data inherently as a result of the nature of the work being done. A short paragraph from the summary section of the funding application should suffice. This section is intended to give a high level overview of your methodology

and data collection procedures. It should provide a standalone description of your project which sits outside your actual funding application.

> The materials employed in these experiments are quite different from any that have been used to study interference. In former studies the subjects were given practice in responding to a set of stimuli until associative bonds were formed between the stimuli and the desired responses, then a change was made in the experimental 'set up' which demanded a different set of responses to the same set of stimuli. In the present study pairs of conflicting stimuli, both being inherent aspects of the same symbols, are presented simultaneously (a name of one color printed in the ink of another color -- a word stimulus and a color stimulus). These stimuli are varied in such a manner as to maintain the potency of their interference effect. In addition we intend to test a cohort of adults as compared with previous studies in the area which worked with children and your adolescents.
>
> Employing new materials and paradigm necessitates that new data are captured from participants.

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

Within the main text of the funding application it is likely there is a short paragraph that covers methodology. Typically this would give enough detail to provide an overview of how data is created. For example, it may be that participants generate data in response to stimuli that are shown on a computer screen, or in a clinical setting a persons conversations with a clinician might be digitally recorded. In this section you a free to go into more detail if relevant. Additional detail may include file creation and naming conventions. For example, each participant will generate a single electronic data file for each experiment they take part in. Each file will automatically be numbered sequentially by the version of E-prime to be used to administer experiments and collect data.

> Participants will take part in a series of experiments in which they will be presented with various coloured words on a computer screen. These words will be presented using E-Prime. Each participant will automatically be assigned a unique number by the E-Prime software for each experiment they participate in. After each experiment data will be stored in a file where the participant number forms the filename and the extension will be ".dat".
>
> Collected data will be combined and analysed using E-Prime E-DataAid which generates ".edat" files. Further analysis will be carried out using SPSS which generates ".sav" files when storing data and ".spv" or ".spo" files when storing the output of statistical tests.
>
> Files will be stored under a meaningful folder structure representing the experiment carried out and each file will have a relevant file name.

2.3.3. Which file formats will you use, and why?

In practice there are only a limited number of file formats and types that most researchers in the psychological science might use. It is accepted that specialised or bespoke hardware and software may generate unique formats. Where possible open source file formats should be chosen in order to decrease the risk of obsolescence and increase open access, i.e. wider reuse amongst the research community. If this is not possible a mirror file, in appropriate format, should be created where possible. For example, alongside a Microsoft Word Document, it may be appropriate to create a pure text, ".txt", file. This rule of thumb could also apply to data files where SPSS ".sav" files are also mirrored by a Comma Separated Value ".csv" file. In short the mirror file provides access to your data using a more widely used and less specialised format in the sense that more software can read those file types.

You should list the types of file you intend to use, along with a reference to the software you will or might use to create it and a one sentence reason why you intend to use that format of file. Typically you will be using software provided by your host institution. Unless you have a specific reason for using a given file format or application the former reason will usually suffice.

During the course of the research we envisage that the following file formats will be used:

File format	Software used	Reason for use
*.docx	Microsoft Word (http://office.microsoft.com/en-gb/)	Standard file format of Microsoft Word which is the word processor application provided by default by my host institution
*.xlsx	Microsoft Excel (http://office.microsoft.com/en-gb/)	Standard file format of Microsoft Excel which is the spreadsheet application provided by default by my host institution
*.sav	IBM SPSS (http://www.spss.com/)	Standard file format of SPSS which is the statistical analysis application provided by default by my host institution
*.spo & spv	IBM SPSS (http://www.spss.com/)	Standard file format of SPSS which is the statistical analysis application provided by default by my host institution
*.es	E-Prime (http://www.pstnet.com/)	Script file that E-Prime generates when designing an experiment
*.edat	E-Prime (http://www.pstnet.com/)	Data file that E-Prime generates when saving a participants responses to trials in an experiment
*.txt	Raw text file	A raw text version of any word processed documents to aid in preservation and open access
*.csv	Comma Separated Value file	A raw text version of data files
*.pdf	Adobe Acrobat (http://www.adobe.com/)	A portable document format version of all files will be produced to aid in preservation and open access

2.3.4. What criteria will you use for Quality Assurance/Management?

As a matter of course you should state how experiments, statistical analyses or clinical work will produce and maintain data that is valid, reliable and accurate. This can include mechanisms such as peer review, calibration of equipment, consistent data format etc.

As a matter of course equipment used to collect timing critical data will be calibrated using a Black Box ToolKit (http://www.blackboxtoolkit.co.uk/) to ensure that data collected is valid, reliable and accurate. All experimental paradigms will be administered through E-Prime which ensures that data will be collected and stored in a consistent format. Data analysis will be carried out using E-DataAid, Excel and SPSS. Where possible metadata will be saved along side raw data files, e.g. from the SPSS data dictionary.

Data generated, and any subsequent analyses, will be internally peer reviewed by members of the project team as appropriate. Any data subsequently published will be subject to wider peer review.

2.4. Relationship between old and new data

2.4.1. What is the relationship between new dataset(s) and existing data?

It may be that the proposed study is a direct replication of previous work and that all data map exactly in terms of experimental procedure, variable specifications and statistical analyses etc. This could be the case irrespective of investigator should the previous studies data be freely available for reuse. If this is the case in this section you can explain the relationship. For example, "Data collected in the proposed study will be identical in format to, XXX project (2009): AN Author, funded by ESRC and based at Any Town Institution". If there is a modification in the methodology that makes the data slightly different in some respect you should outline the potential differences.

There is no relationship between existing data and the proposed study by virtue of the use of novel materials and new procedure as outlined in 2.3.1.

2.4.2. How will you manage integration between the data being gathered in the project and pre-existing data sources?

If you intend to reuse existing data as part of the proposed study you should outline the way you plan to integrate old and new data. This should include technical aspects, e.g. "We plan to merge datasets using E-Merge (E-Prime) and SPSS"; You should also include a short statement on data provenance in which you should state where and how the data was produced, who has modified it and how old it is along with any other pertinent information, e.g. "Pre-existing data was generated by the current principal investigator as part of the XXX project (2009): AN Author, funded by ESRC and based at Any Town Institution and has trusted provenance."; Finally you should include a short statement on data quality, e.g. "The quality of data generated as part of the XXX project (2009): AN Author, funded by ESRC and based at Any Town Institution is known and raw unmanipulated datasets are available."

There is no relationship between existing data and the proposed study by virtue of the use of novel materials and new procedure as outlined in 2.3.1.

2.4.3. What added value will the new data provide to existing datasets?

Where studies collect data that will be added to, or merged with, an existing corpus you should succinctly outline the value which this new data provides. For example, "By collecting data from a different cohort this will make predictions based on the larger combined dataset more robust and more widely applicable." If there is no relationship to existing datasets as a result in differing experimental methodology for example you can succinctly mention it here.

There is no relationship between existing data and the proposed study by virtue of the use of novel materials and new procedure as outlined in 2.3.1.

2.5. Data Documentation and Metadata

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

Generally research datasets are not understandable in isolation without additional explanatory documentation. For example, data that was held in SPSS data files until recently did not support long variable names. This meant that anyone working with such files had to keep a separate record of what variable names represent. In more recent versions of SPSS long variable names can be used together with a data dictionary to enhance clarity. Without such mapping and explanatory notes you should answer "no". If you answer "yes", you should outline how the data are understandable in isolation, e.g. "Yes. Datasets are understandable in isolation as each is stored in SPSS format data files and is described via long variable names and inbuilt data dictionary. Detailed metadata notes are also provided as to what each variable represents." It is possible that your data may be qualitative in nature. Again without metadata or descriptive text this would not be understandable in isolation. For qualitative data, software such as NVivo for example, may be able to generate appropriate explanatory metadata automatically.

No.

2.5.2. If No, what contextual details are needed to make the data you capture or collect meaningful?

In order to make data captured in an experimental setting meaningful to others, or yourself at a later date, it is useful to create metadata. Metadata basically describes, or annotates, your data to give it contextual information. JISC describe metadata as follows:

> "Metadata is often defined as 'data about data' or 'information about information'. In the digital world, metadata is usually structured textual information that describes something about the creation, content, or context of an individual file or collection of many digital files.
>
> Metadata might take the form of controlled terminology, carefully constructed or chosen from formal lists and entered into pre-established categories. Or it might be simply a free text description or set of keywords used to annotate or 'tag' an image. It might describe something objective and straightforward, such as the file size of the digital file; or something much more complex, such as the subject matter of the resource or legal rights associated with its use. Metadata is often held within databases, but it can take other forms - it can just as easily be found embedded within the digital file itself. In short, metadata provides the means for us to describe our digital resources in a structured way that enables us to share those resources with other people and machines."
>
> http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/an-introduction-to-metadata/

Constructing metadata is generally a task that can only easily be done by the principal investigator. Ideally it should be constructed alongside the data in the generation phase and where possible standardised tools should be used. SPSS for example can aid the process by utilising its data dictionary to generate metadata in XML format. This type of metadata is known as intrinsic or implicit metadata. On the less machine orientated side text and/or Excel files describing datasets narratively are equally effective and should be used to provide as much context as possible to data, materials and procedures. This type of metadata is called extrinsic or explicit metadata and is generally the most useful for others wishing to access shared data. In short you should explain what variables represent, whether they hold independent or dependant variables, what ranges they cover and so on. For data that is not quantitative in nature it is still possible to describe that data narratively so that others can understand the structure of the data and make use of it at a later date.

Datasets stored in SPSS will be fully described using meaningful long variable names together with entries in the data dictionary. Additional extrinsic, or contextual, information about the dataset will be stored in either text files and/or Excel spreadsheets. For example, each SPSS data file will have an accompanying text file that will describe the dataset, variables, variable scope etc. together with annotated examples to aid clarity.

2.5.3. How will you create or capture these metadata?

Ideally intrinsic metadata should be automatically generated where possible by whatever software packages you are using to store, analyse or process your data. For example, SPSS can generate metadata for variable names, data types, lengths, and coding from variable names and its data dictionary. However, this is dependent on you carefully considering how you name and describe variables to begin with.

Generally there will be a mix of intrinsic metadata, which is automatically generated, and extrinsic metadata manually crafted by the principal investigator or other members of the research team. Often it is extrinsic metadata which is of most use to other researchers and to the investigator themselves in the future. A good analogy is the pocketbooks that researchers of yesteryear used to note down aspects of experimental procedure as they went about their business. These could be referenced easily to aid in the replication of previous work. Replication should be at the heart of good science and therefore so should keeping detailed metadata.

Datasets stored in SPSS will be fully described using meaningful long variable names together with entries in the data dictionary. An XML file will be generated automatically from these in order to intrinsically describe the dataset. Additional extrinsic, or contextual, information about the dataset will be stored in either text files and/or Excel spreadsheets. This extrinsic information will be generated by the principal investigator.

2.5.4. What form will the metadata take?

Metadata is made up of a number of elements which can be categorised into the different functions they support. A metadata standard will normally support a number of defined functions, and will specify elements which make these possible. A metadata standard may support some or all of the following functions:

- Descriptive Metadata enables identification, location and retrieval of information resources by users, often including the use of controlled vocabularies for classification and indexing and links to related resources. In the psychological sciences, these may be weblinks to related data files, experimental scripts and so on.
- Technical Metadata describes the technical processes used to produce, or required to use a digital object. These would typically be the metadata and contextual information provided in response to the previous questions.
- Administrative Metadata is used to manage administrative aspects of the digital object such as intellectual property rights and acquisition. Administrative Metadata also documents information concerning the creation, alteration and version control of the metadata itself. This is sometimes known as meta-metadata! This would typically describe how you access the actual datasets and accompanying descriptive metadata. It may provide contact names, phone numbers, emails and acceptable use policies etc. It may also include information on copyright and licencing terms, e.g. Creative Commons licence details.
- Use Metadata manages user access, user tracking and multi-versioning information. For example it may include details of previous studies in which that data was used along with version numbers highlighting chronological differences between data files.
- Preservation Metadata, amongst other things, documents actions which have been undertaken to preserve a digital resource such as migrations and checksum calculations. For example, this kind of metadata explains what was chosen for preservation and the rationale for selection.

http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards

Generally the researcher will only be concerned with two or three forms of metadata. These will encompass Technical Metadata which includes intrinsic metadata, e.g. information about data files, such as file specifications and data dictionaries generated automatically by SPSS, and extrinsic metadata which gives contextual information about how and why the data were generated and what they represent, e.g. how to make use of the data. It is possible that Administrative Metadata may also be generated for either the whole project, or individual datasets, in order to cover issues such as intellectual property rights. Finally Descriptive Metadata can be constructed to describe the gross features of the dataset, i.e. to describe what the dataset contains. This is typically done so that search engines can index the data and make it discoverable, e.g. through Google. A good standard for this is the widely recognised Dublin Core (ISO 15836:2009) which consists of 15 simple metadata elements covering data such as title, author names, location of resource etc. Conceptually this is akin to a library catalogue record. For more details on the 15 elements see: http://dublincore.org/documents/dces/

Descriptive metadata will be created using the Dublin Core standard (ISO 15836:2009) which will allow other researchers and automated tools, such as search engines, to discover and catalogue the dataset according to recognised standards. Dublin Core metadata will be stored as XML 1.0 records as appropriate.

Extrinsic descriptive metadata will be created in XML 1.0. This will narratively describe the dataset and its features to enable reuse and repurposing by others. The structure of this metadata will be appropriate to the project and the judgement of the principal investigator.

Intrinsic technical metadata will be automatically generated in XML 1.0 from the SPSS data dictionary in order to describe the datasets attributes. The structure for this metadata will be constrained by the support offered in the current release of SPSS.

2.5.5. Why have you chosen particular standards and approaches for metadata and contextual documentation?

You should choose metadata standards based on ease of use and how well adopted they are. For example, you should not choose seldom used metadata standards or ones that are complex to implement. For example Dublin Core's 15 elements are easy to understand, simple to implement and are well used. Often your options for intrinsic metadata may be restricted in terms of the formats and structure the software you are using can produce, e.g. the format and structure of the XML metadata that SPSS can generate in order to describe its data dictionary. For a brief introduction to XML see http://en.wikipedia.org/wiki/XML.

Often contextual metadata will need to be generated by the principal investigator themselves as they are best placed to do so. Again XML files are the recommended format of choice as they offer a way of semantically coding contextual information. Think of XML files as a text document with structured section headings, where the researcher can define the headings and the contextual information to put under them. Various editors allow such files to be constructed easily where elements can simply be thought of as section headings.

> "XML defines rules to mark-up a document in a way that allows the author to express semantic meaning in the mark-up. XML does not necessarily restrict the author to certain tags (elements) as HTML and XHTML does."
>
> From http://icant.co.uk/webstandardsforbusiness/pmwiki.php/Main/MACCAWS-Glossary

Dublin Core (ISO 15836:2009) was chosen as the descriptive metadata format of choice as it is a widely adopted format and simple to implement in XML 1.0. XML 1.0 documents will be used to store all metadata as it is a widely adopted format, simple to implement, cross platform and machine readable. Contextual information will be stored in XML 1.0 format documents as a result. Intrinsic metadata will also be stored in this format as SPSS can automatically output XML 1.0 from data files data dictionaries. Where appropriate supporting text and/or Excel files will provide a narrative explanation of the dataset and its contents. The data itself will be stored in its natve format alongside a simpler mirror format, e.g. SPSS *.sav files and Comma Separated Value (CSV) files.

3. Ethics and Intellectual Property

3.1. Ethical and privacy issues

3.1.1. Are there ethical and privacy issues that may prohibit data sharing?

Generally there will nearly always be ethical or privacy issues that initially prohibit wider data sharing. For example, often individuals who have taken part in an experiment, or clinical study, will be personally identifiable along with their data by their name or date of birth. If this is the case you should answer "yes". If you definitively have no personally identifiable data then you should answer "no".

Yes.

3.1.2. If Yes, How will these be resolved?

In most standard experimental paradigms participants are identified by an anonymised number or in some cases have no identification at all other than something identifying which experimental condition they were in. E-Prime for example identifies subjects using an incrementing number meaning that if you want to tie an individual back to an experiment you will need an external mapping of some sort. If your datasets contain personally identifiable information such as names, dates of births, addresses, NHS numbers etc. it is possible to remove these to enable sharing of anonymised data. It may be that for specific types of data and scenarios that referral to departmental or institutional ethics committees is required. Formal consent of the individual(s) concerned may also be appropriate.

Any datasets that contain personally identifiable information will be anonymised according to standard best practice. As E-Prime is being used to administer all experimental paradigms we don't envisage a large volume of anonymisation will be necessary as participants data is automatically anonymised and output file names sequentially numbered rather than use any personal biographical data or experimental condition information.

3.1.3. Is the data that you will be capturing/creating "personal data" in terms of the Data Protection Act (1998)?

For more information on the Data Protection Act 1998 you should consult a specialist within your host institution. There will be a named person who will be responsible for and registered with the Information Commissioner's Office.

Personal data includes all biographical data for example, e.g. name, addresses, National Insurance numbers, medical records. In short anything that identifies an individual directly or indirectly.

In general scientific research data can be anonymised in order to ensure that any data is not "personal data". In clinical settings this may not be an option. In such instances you are advised to consult a local expert or if working within an NHS trust, the Caldicott Guardian.

In short if you answer “no” there should be no way of linking data back to a individual whether this be directly or indirectly. Indirectly means by using conditional information or data such as Date of Birth to deduce and identify indviduals.

No, as all data will be anonymised and at no stage will any identifying personal data be held.

3.1.4. What action will you take to comply with your obligation under the Data Protection Act (1998)?

In order to comply with the Data Protection Act 1998 it may be sufficient to state eight key points outlined within the act itself. For more information on the Data Protection Act 1998 you should consult a specialist within your host institution. There will be a named person who will be responsible for and registered with the Information Commissioner's Office. It is possible that your institution has a standard set of procedures to help ensure you comply with the act.

The project will comply with its obligations under the Data Protection Act 1998 by following the eight principles outlined below:

1. Fairly and lawfully processed, in particular in accordance with certain conditions in the Act
2. Obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes
3. Adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed
4. Accurate and, where necessary, kept up to date
5. Not kept longer than necessary for the purpose or purposes they are processed for
6. Processed in accordance with the individual's rights under the DPA
7. Kept secure by the taking of appropriate technical and organisational measures against unauthorised or unlawful processing and accidental loss, destruction or damage
8. Not transferred to countries outside the European Economic Area, unless there is adequate protection

3.2. Intellectual property rights

3.2.1. Will the dataset(s) be covered by copyright or the Database Right?

Generally anything which is generated by a project is automatically copyrighted and ultimately assigned to either the host institution and/or the funding body. So the answer in most cases here will be, "Yes". You are strongly advised to contact the Research Support Office within your institution or the funding body to clarify the position.

Yes.

3.2.2. If Yes, Who owns the copyright and other Intellectual Property?

Generally copyright will be owned by the host institution and the funding body. However you are strongly advised to contact the Research Support Office within your institution or the funding body to clarify the position.

Copyright will be shared between the host institution (The University of York) and ESRC.

3.2.3. If Yes, How will the dataset be licensed?

Research councils are increasingly requiring that any data and other project outputs be shared via Creative Commons licences. If this is the case it will be made explicit in their funding guidelines along with the type of licence you should be releasing data and other project outputs under.

For example, the "Attribution Non-Commercial Share Alike" agreement may be appropriate:

Attribution (BY) Non-Commercial (NC) Share Alike (SA) otherwise known as a “cc by-nc-sa” licence. This license lets others remix, tweak, and build upon your work non-commercially (NC), as long as they credit you (BY) and license their new creations under the identical licence terms (SA).

Licencing is used to put restrictions on what others can do with the outputs of your work, e.g. the datasets you produce and have chosen to share with the wider community. For further clarification you are advised to contact your funder and/or institutions Research Support Office. You can also find out more about the types of Creative Commons licences by visiting http://www.creativecommons.org/. Creative Commons licences are widespread in other domains, e.g. images on Flickr.com or videos on YouTube.

Any datasets and other project output will be licensed under the Creative Commons "Attribution Non-Commercial Share Alike" (cc by-nc-sa) agreement.

3.2.4. What is the dispute resolution process/mechanism for mediation?

It is possible that once you share your data there may be a breach of the licensing conditions which you choose to release your data, or other project outputs, under. For example, if you released your data under the Creative Commons "Attribution Non-Commercial Share Alike" (cc by-nc-sa) licence and a third party choose to reuse your data for commercial purposes this in would in theory create a dispute. If a dispute arises you are advised to seek expert advice and to pass the matter onto those who routinely deal with such disputes. This may be a department within your institution, e.g. the Research Support Office, a legal team at your funders, or if it occurs after the project finishes the repository where you chose to deposit, e.g. the UK Data Archive. Any of these bodies should have a recognised dispute resolution process and mechanism for mediation. As the exact resolution processes and mechanisms for mediation are likely to vary according to the body involved and the nature of the dispute you are advised to try and list your various points of contact.

Any licensing disputes will be passed onto the relevant body depending on the nature of the dispute and where it occurs in the lifecycle of the project. This could be any or all of the following, the Research Support Office at The University of York, ESRC, and UK Data archive. Each body has a defined dispute resolution process and mechanisms for mediation.

4. Access, Data Sharing and Reuse

4.1. Access and Data Sharing

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

Some funding bodies specify that some or all project outputs be shared within the wider community to promote reuse and repurposing. This may be a funding requirement and if so the obligation to do so will be clearly articulated within the funding guidelines. If so answer yes.

Yes.

4.1.2. If No, why will you not share your data?

Certain projects will be barred from sharing data as a requirement of the funding or other criteria, e.g. defence grants, medical research etc. If you answered "no" to question 4.1.1 you should enter the reason here. In most cases this will be clearly articulated within the funding guidelines and can be restated here.

4.1.3. If Yes, How will you make the data available?

Data sharing can be onerous for individual researchers if they have to deal with each request for data personally. However services exist that will archive data and handle the mechanisms by which it is shared. For example, Jorum (JISC) and The UK Data Archive (ESRC) are two such services. It may also be the case that depositing data in such a service is a requirement of funding. If this is a requirement this will be clearly articulated within the funding guidelines and can be restated here.

As a requirement of funding all datasets will be deposited in The UK Data Archive as outlined in the ESRC funding guidelines. Data will be made available through The UK Data Archive standard operating procedures.

4.1.4. If Yes, When will you make the data available?

In most cases if data is to be shared it will be made available at the end of the project. In some cases it may be that there is a phased release of data throughout the project. In rare cases data may be embargoed for a set period either as a requirement of funding or due to the nature of the project, e.g. commercially exploitable outputs. You should succinctly state approximately when data will be made available.

Data will be made available at the end of the project.

4.1.5. If Yes, What is the process for gaining access to the data?

All projects are advised to deposit data in a data repository and sharing service such as The Data Archive. As a result there will be a clearly defined process to access the data which can be succinctly stated here.

Data will be deposited with The Data Archive on completion of the project. In order to access data prospective users will need to contact The UK Data Archive (http://www.data-archive.ac.uk) and follow standard operating procedures.

4.1.6. If Yes, Will access be chargeable?

If you are depositing with a recognised repository there may be a set charging policy in place or it may operate without charge. Here you should state what the current policy is for gaining access to datasets such as yours. If you are unsure you should contact the data repository service you intend to use for guidance. For example, The UK Data Archive may charge a nominal fee for access to certain types of dataset.

Yes.

4.1.7. If Yes, Please give details

If access to data stored within your chosen data repository service is chargeable you should give details here. You only need highlight that data access may be chargeable and that potential users should refer to the repository service itself for more details.

The Data Archive currently charge a nominal fee for access to certain types of dataset. For current details of pricing you should consult The UK Data Archive.

4.1.8. Which groups or organisations are likely to be interested in the data that you will create/capture?

Often it will be researchers in the same or cognate fields to those involved in the current project that may want to access the data at some future point. It may be difficult to precisely indicate who these future users might be outside those outlined. You do not need to offer a definitive list but rather provide a general indiction.

Researchers interested in cognitive psychology and specifically perception.

4.1.9. How do you anticipate your new data being reused?

Again it is difficult to predict how others might want to reuse the current projects data in the future. You are advised to briefly state possible areas of reuse.

It is possible that others may want to reuse this project's data in order to generate new hypotheses to explain current findings or as a basis to construct new theories and areas of research. Alternatively they may be used for instruction within a teaching context.

4.2. Exploitation

4.2.1. Does the original data collector/creator/principal investigator have the right to use the data before opening it up to wider use?

It may be that the current project, as a requirement of funding or due to commercial exploitation reasons, has a right to exclusively use the data before releasing it for use within the wider community. Here you should state whether this is the case. For example, patents on some aspect of the project may exist that allow for commercial exploitation for a given length of time before data need be released. In addition the researchers who created the original data also have a timeframe in which they are allowed exclusive access to the data in order to publish ahead of other researchers once the data has been shared more widely.

Yes.

4.2.2. If Yes, Please give details

If you answered "yes" to question 4.2.1 you should give succinct details here. For example, the funder may allow the researcher to exploit the outputs of the project for a fixed duration before making datasets freely available. You should check with both your institution and funders for further guidance.

Standard terms of ESRC grant funding allows researchers exclusive access to data they generate during a projects lifetime and for a further period of six months after the completion date.

4.2.3. Are there any embargo periods for political/commercial/patent reasons?

It may be that the reasons for not releasing data for wider reuse immediately may be due to embargo periods as a result of political/commercial/patent reasons. For example, pharmaceutical research may be protected by patents for a number of years.

No.

4.2.4. If Yes, Please give details

If there any embargo periods as a result of political/commercial/patent reasons you should detail them here if you answered "yes" to question 4.2.3.

4.3. Publications

4.3.1. Do you plan to publish findings which rely on the data?

Most projects will wish to publish findings which rely on the data collected either as traditional journal articles or via other avenues. It may be a funding requirement that you disseminate findings through appropriate methods. Generally there will be a dissemination section within your funding application.

Yes.

4.3.2. If Yes, do your prospective publishers place any restrictions on other avenues of publication?

Most publishers place some restrictions on publishing identical or similar articles simultaneously with other publishers or journals. However this generally applies to textual content rather than the data itself. If there is some restriction you should answer "yes" and then explain what this is in question 4.3.3.

No.

4.3.3. If Yes, Please give details

If your prospective publishers are likely to place a restriction on other avenues of publication you should succinctly outline it here. For example, they may request that identical articles are not submitted for review or simultaneously published elsewhere.

5. Short-Term Storage and Data Management

5.1. Short-Term Storage Requirements

Here you should attempt to quantify the volume of data that you will create over the short term. These figures do not need to be exact and are only an indication volumes. You are free to express volumes in a format relevant to your project. This could be in terms of completed paper questionnaires, Dictaphone tapes or GigaBytes of data if you are storing data as computer files. You may need specialist advice from your department in order to estimate this.

> Over the course of the study we anticipate a dataset of 500Mb of data will be generated each month on average.

5.2. Storage Media

5.2.1. Where (physically) will you store the data during the project's lifetime?

In most cases data will be stored on a computer's hard drive within laptop, desktop, server or external enclosure of some sort. Data may also be held in "the cloud" and may not have a physical location as such. It is also possible that data may be held on audio tapes in clinical settings or even on paper within filing cabinets for questionnaire responses. Here you should state where the data will be physically held. It is possible that data may be held in more than one location if it is physically mirrored or if different data is collected at multiple sites.

> Data will physically be held on desktop PCs hard drives. The desktop PC will be located in a secure location in the named institutions relevant departments.

5.2.2. What media will you use for primary storage during the project's lifetime?

You should state the type of data storage media that you intend to use during the project's lifetime. This encompasses not just computer based technologies but also any audio tape or printed paper. Preferably project data will be stored on hard disk which will be part of a secure department fileserver, i.e. a mapped or shared drive letter. Alternatively data could be held on the internal hard disk of a desktop PC or laptop.

> The primary data storage media during the project's lifetime is likely to be a hard disk drive which is located within the departments secure fileserver.

5.2.3. How will you transfer/transmit the data, if required?

During the lifetime of a project it may be necessary to transmit data between project partners, across sites, upload it to funders or disseminate it to others via a website. This could also encompass sending data between sites manually, e.g. on audio tapes, printed paper or CDs etc. Any data you transmit should be sent securely and/or passworded and encrypted. You should mention both the medium and the security you intend to use. For example, Virtual Private Networks, or VPNs, allow your computer to see a shared drive letter on another computer or server anywhere in the world over an encrypted connection.

> If required, data will be transferred between sites using Virtual Private Networks (VPNs) encrypted to recognised standards as implemented at each institution by central services.

5.3. Back-Up

5.3.1. How will you back-up the data during the project's lifetime?

Many institutions and departments will have mature and reliable backup mechanisms, or policies, in place which can be utilised during the project's lifetime for data stored on computer. This means that the intricacies of a backup strategy need not concern the individual investigator. If your institution has a central backup strategy, or policy, in place a copy of the policy can be included here or a reference to it. If no institutional or departmental mechanisms exist then you should say how you intend to regularly backup data. In reality it is useful to not just rely on global backup

mechanisms or policies and to also backup at the level of individual investigator. For none computer stored data you may also wish to consider a backup strategy, e.g. photocopying paper based questionnaire responses and storing copies at another off-site secure location, or duplicating Dictaphone tapes used in clinical settings.

> Global institutional and departmental level backup mechanisms and policies exist which cover the project during its lifetime. Additionally data will be backed-up to an external hard drive. A second encrypted external hard drive will be taken off-site and stored in a secure location.

5.3.2. How regularly will back-ups be made?

Once you have outlined your backup mechanism or policy you should state how frequently backups will be made. For institutional or departmental mechanisms and policies this will already be determined. If you are managing your own backup strategy you should detail the frequency and methodology here. It is beyond the scope of this guide to outline a personal backup strategy and as such you are advised to seek advice within your local department.

> Global institutional and departmental level backup mechanisms and policies exist which determine the regularity of backups according to best practice. Additionally data will be backed-up to an external hard drive on a nightly basis. A second encrypted external hard drive will be taken off-site at the end of each month and stored in a secure location.

5.4. Security

5.4.1. How will you manage internal access restrictions and data security during the project's lifetime?

Most computers require a username and password to logon to both the local computer and network. In some settings, such as NHS trusts, this may be supplanted with swipe cards or biometric mechanisms such as fingerprint recognition. Again institutional or departmental guidelines may exist covering password strength and the frequency that passwords must be changed. These processes may be automatically enforced at a local level. If you manage your own password protection you should state how you intend to maintain password strength, e.g. length, inclusion of letters and numbers and symbols, together with how often you intend to change it. You may also want to mention encryption of data at this point if it is tied into a local security mechanism, e.g. username and password, swipe card or biometric security. If you are working with data held on Dictaphone tapes or questionnaire responses held on paper then you should outline how these are securely held, e.g. within a locked filing cabinet, within a locked room etc.

> Global institutional and departmental password mechanisms and policies exist which determine the strength of and frequency of changes according to best practice. Data held on local and networked storage will be encrypted based on these credentials.

5.4.2. How will you manage permissions, restrictions and embargoes?

It is possible that various investigators involved in the project will need varying levels of access to different data or other resources. That is, a principal investigator may be able to see all data whereas a research assistant may only be able to see data or files they have been given permission to view. You should succinctly state how varying levels of access are maintained or granted. Again this may be managed at a global institutional or departmental level.

> Access to data and files stored on the departmental fileserver will be restricted by staff seniority and maintained though Active Directory and local Windows security.

5.4.3. Other security issues

Other security issues can encompass the security of sensitive data and whether data is taken off-site on laptops, smartphones or memory sticks for example. It can also encompass policies on copying data. Other security issues can encompass requirements needed to meet the Caldicott guidelines in an NHS trust for storage of clinical data.

> No data will be taken off-site apart from the monthly hard drive based backup which will be encrypted and ultimately stored in a secure location.

5.4.4. Is it appropriate to transfer this data across unsecured network connections?

Any data should only be sent across an unsecured network where that data is not of a sensitive nature or includes personally identifiable information that would necessitate it be covered by the Data Protection Act 1998.

Data that is not subject to the Data Protection Act 1998 may be transmitted over unsecured networks where an encrypted alternative does not exist.

6. Deposit and Long-Term Preservation

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

Many projects will deposit data into a recognised repository at either a departmental, institutional or national level. Within institutions there is likely to be a archivist or records manager who will be able to offer advice on where to deposit data for safe and secure long term storage. Some funders will also place requirements on where data should be deposited. For example, ESRC requires that funded projects, on completion, lodge data with the UK Data Archive (http://www.data-archive.ac.uk) which it part funds for this purpose. If you are working within an NHS trust you should consult your local Caldicott Guardian.

Generally researchers are advised not to try to maintain, curate and archive data either personally or at a project level in isolation, i.e. as the primary source.

> On completion, project data will be deposited with The UK Data Archive as a requirement of ESRC funding. The UK Data Archive acquire, curate and provide access to the UK's largest collection of social and economic data.

6.2. Long-Term Specifics

6.2.1 Selection: What data will be preserved for the long-term?

In most cases you should store all "meaningful" data where practicable. Meaningful in the sense that the data has been collected from genuine participants and does not represent test data of any kind. You are advised to store raw data along with any data which is derived as a result of data manipulation, e.g. sampling or filtering data.

If you manipulate raw data in any way in order to produce derived data you are advised to document such transformations in a metadata file in XML or plain text that outlines both the process and logic behind it. This descriptive file should be stored with both the raw data, which itself should be suitably described, and with the derived data.

> All data produced as a direct result of the project will be preserved along with any data derived from that data.

6.2.2. Selection: On what basis will data be selected for long-term preservation?

If you have chosen only to preserve a selection of your data you should outline your selection criteria here. In most cases all your data will be preserved as a matter of course.

> Does not apply as all data will be preserved.

6.2.3. Retention: How long will/should data be kept beyond the life of the project?

Generally data retention length will be dependent on guidance from the host institution or funding body. For example, some institutions recommend research data should be held for four years after the end of the project if unfunded and seven years if funded. If data is of historical value they may recommend it is stored with a national subject based repository (http://www.kcl.ac.uk/content/1/c6/05/57/93/Howlongshouldikeepmyresearchdata.pdf). It may be the case that if dealing with certain kinds of NHS data it should be held for 20 years or more (http://www.nhs.uk/chq/Pages/1889.aspx?CategoryID=68&SubCategoryID=160). If you are working with any kind of NHS data you local Caldicott Guardian should be able to offer advice.

Political, temporal, commercial and legal issues may also play a role. One should also remember that it may be hard to objectively determine if data is worthy of storage for historical reasons if the research has only recently been completed.

If you can succinctly outline the length of time the data will be held beyond the life of the project you should do so here. Alternatively you can provide a link to the institutional, chosen repository or funding bodies policy.

All data, as a funding requirement, will be deposited with The UK Data Archive. Data deposited with The Data Archive will be stored for the number of years outlined in their data storage policy (http://www.data-archive.ac.uk).

6.2.4. If the dataset includes sensitive data, how will you manage this over the longer term?

If the dataset includes sensitive data you may need to include provision for management of that data over the longer term, e.g. to allow for deletions. In most cases this will not apply to traditional research data which are in the main anonymised.

If you are storing sensitive data over the long term you are advised to consult your institution, chosen repository or funder for further guidance. If dealing with NHS data this may involve liaising with you local Caldicott Guardian.

The dataset includes no sensitive data that needs to be managed over the long term.

6.2.5. Transformation: Will transformations will be necessary to prepare data for preservation/data sharing?

It many cases before data can preserved it will need to be transformed in some way. This can include cleaning the data or anonymising it so that any personal information is removed or it may involve saving the data into a more widely used format. In some instances this may be a specialised task that needs to be carried out by the repository into which you intend to deposit your data.

Yes

6.2.6. Transformation: If Yes, what transformations will be necessary to prepare data for preservation/data sharing?

If transformations are necessary you should state which ones will be applied before data can be preserved and/or shared.

All data will be anonymised to remove any personally identifiable information before it is deposited with The UK Data Archive and shared with the wider community as is a requirement of ESRC funding

6.3. Metadata and Documentation for Long-Term Preservation

6.3.1. What metadata/documentation will be submitted alongside the datasets or created on deposit/transformation in order to make the datasets reusable?

Generally you will need to store some form of metadata or explanatory documentation along with your data otherwise it will be difficult to contextualise. Where possible contextualising documentation should include the original bid, research reports and publications. Further metadata should describe both the actual data and other administrative information. You should also include details of any manipulations carried out on the data. You should construct this information in the mindset of yourself wishing to reuse the data in say 10 years time or from the prospective of someone who has not worked intimately on the project but wishes to reuse the data.

You are advised to consult your local archiving specialist or intended repository for further advice on what supporting information is suitable and the depth that this needs to go in to. It may be that your intended repository will aid you in the creation of metadata through specialised forms or will help create it for you.

Contextualising information such as the original bid, research reports and publications will be stored alongside the projects data outputs. Metadata will be constructed to describe the actual data itself together with relevant administrative information to aid in resource discovery. If any manipulations have been carried out these will be described in suitable depth.

6.3.2. How will this be created, and by whom?

Metadata and supporting documentation is often best created by those intimately involved in the project. That is not to say that expert advice on the types of information that may be required should not be sought. Often the repository where the data will be deposited will be best placed to offer such advice.

Metadata and supporting documentation will be created by the Principal Investigator under the guidance of The UK Data Archive.

6.3.3. Will you include links to published materials and/or outcomes?

Projects will produce various reports, research papers and web outputs during their lifetime which can be linked to and so the answer will invariably be "yes".

Yes.

6.3.4. If Yes, please give details

Often projects will have a web presence which can form the first point of contact for a project. Ideally you should provide a link to this and, where possible, to the funders web pages where your project is listed. If you have produced reports, research papers and other web outputs during the lifetime of the project you can also include links to them directly if hosted on third party servers, e.g. journal websites, citation indexes etc.

Web links to the main project website will be provided in order to allow access to contextual information, reports, research papers and other outputs. Where project outputs are hosted by third parties, such as journal publishers, direct links will be provided.

6.3.5. How will you address the issue of persistent citation?

Many project's end resources tend to remain available for only a relatively short time if they are self-hosted. As a consequence copies of documents can become spread around the internet which can have the unwanted side effect of introducing multiple versions, or drafts, of the same document rather than one definitive version.

One way round this issue is to have persistent URL's where if the end location of a document or webpage changes a user can still find the document from the same URL. Think of this as a telephone directory where people's names don't change but the numbers they can be reached on can. If people move it's simple a matter of updating the number rather than how you find their entry.

The Digital Object Identifier (DOI® http://www.doi.org/) System is a form of persistent citation and is used for identifying content objects in the digital environment. DOI® names are assigned to any entity for use on digital networks. They are used to provide current information, including where they (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI name will not change.

DOI's generally look like this: "10.1000/182" and can be used to find the actual resource. Many journal publishers now use them for individual articles and there are various services that can be used to create and resolve them, e.g. CrossRef (http://www.crossref.org/). In order to convert a DOI to a real URL you would visit a resolver service like crossref, enter the DOI, then be directed to the source webpage for the resource, e.g. an academic article.

Often publishers request that both your own article or dataset has one and that any references you cite also have a DOI to uniquely identify them.

Persistent URLs or PURLs (http://purl.org/) are often used for web resources and serve a similar purpose to DOIs, e.g. they may be used to locate a dataset. In much the same way as DOI's they allow for web pages and web resources to be found even when their end location changes.

Where practical you are advised to make use of both DOIs and PURLs

> Persistent citations will be used for both project documents (DOIs through CrossRef) and web resources (PURLs though purl.org) in order to help ensure persistence both during the project and once it has ended.

6.4. Longer-Term Stewardship

6.4.1. Who will have responsibility over time for decisions about the data once the original personnel have gone?

Project funding and life cycles mean that when a project ends personnel often move on and data and other project outputs can be lost. As a result most new projects are advised to deposit their data and other project outputs within a departmental, institutional or nationally recognised repository, e.g. The UK Data Archive in the case of ESRC funding requirements. Typically decisions about the data and other outputs will be devolved to the chosen repository after the project ends. In NHS trusts this responsibility may rest with the local Caldicott Guardian.

> As data and other project outputs will be deposited with The UK Data Archive as a requirement of funding it is anticipated that they will be responsible for longer term stewardship.

6.4.2. In the event of the long-term place of deposit closing, what is the formal process for transferring responsibility for the data?

Repositories can and do come and go over time as a result of funding changes or replacement by other services. The Arts and Humanities Data Service (http://www.ahds.ac.uk/) and Intute (http://www.intute.ac.uk/) are two such examples. For an individual investigator it can be difficult to foresee such issues and control what happens to project data and other outputs. Notwithstanding the best advice is to select a repository that has national standing or has been given as a requirement of funding. Within the NHS issues such as closure will be dealt with centrally. By choosing a national archive one can devolve issues such as this to them as they should have suitable policies in place.

> In the event of The UK Data archive closing the responsibility for transfer of data and other materials is devolved to them.

7. Resourcing

7.1. Outline the staff/organisational roles and responsibilities for implementing this data management plan

You are encouraged to provide an overview of who will do what in order for your data management plan to be implemented. Depending on where you are in the project's lifetime this may only be a sketch outline. Remember it is possible to add more detail as the project progresses. Ideally you should be making use of departmental or institutional provision where possible, e.g. for backing up data and physical security. In NHS trusts this may also involve the local Caldicott Guardian.

In your answer you should draw together the various sections of the proceeding data management plan and succinctly summarise them. For most research projects this will be fairly straightforward.

Role	Responsibility	Time allocation
Daily backup of data	Host department	As required
Daily local backup of data	Principal investigator (J. Ridley Stroop)	As required
Monthly backup of data (offsite)	Principal investigator (J. Ridley Stroop)	As required
Physical security of computer hardware	Host department	As required
Creation of metadata	Principal investigator (J. Ridley Stroop)	1 hour per week (est)
Creation of supporting contextualising documentation	Principal investigator (J. Ridley Stroop)	1 hour per week (est)
Cleaning of data to meet Data Protection Act 1998 requirement, e.g. anonymsing data	Research Assistant (TBA)	1 hour per week (est)
Transposition of data into suitable format for deposit into repository	Research Assistant (TBA)	1 hour per week (est)
Deposit into The UK Data Archive at end of project	Principal investigator (J. Ridley Stroop)	As required

7.2. How will data management activities be funded during the project's lifetime?

Generally data management activities will be funded in two ways during the project's lifetime. If you are making use of departmental or institutional facilities, e.g. fileservers and backup procedures, these will typically be funded through overheads built into the grant. That is, the percentage of the grant that goes to the host institution. Some institutions charge for backup space so you may wish to mention this here if you know yours does. The bulk of the management activities will fall on the personnel working on the project and will be a fraction of their paid time. The project may also purchase equipment where required.

Departmental and institutional shared fileserver space and backup will be funded though host institution overheads included within project funding. The Principal investigator (J. Ridley Stroop) and Research Assistants data management activities will be funded as a fraction of the funding made available.

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

Fully costed longer term management can be difficult to define before a project ends. Some departments and institutions have a policy of charging for the volume of data you wish to store multiplied by the number of years you wish it to remain available. If this is the case you can cost this based on likely funding body requirements, e.g. number of years they wish the data to remain available. Alternatively some host institutions will handle this as a function of institutional

overheads. Longer term storage at a national data archive may be provided at no additional cost to the project as in effect it is met by the funder if they state deposit as a requirement. You are advised to check with your local department, host institution or funder for further guidance.

Data and other project outputs will be deposited with The UK Data Archive as a requirement of funding. Costs for long term data management will be met by ESRC and The UK Data Archive.

8. Adherence and Review

8.1. Adherence

8.1.1. How will adherence to this data management plan be checked or demonstrated?

Checking data management plan tasks against project outputs is a good way of demonstrating adherence. For example, for a study which generates data in a series of experiments it is a good idea to generate contextualising metadata or other descriptive texts in parallel or just after data has been collected. One way of demonstrating adherence to the data management plan is to set a time limit for creating such metadata after a sub-study has taken place. By creating metadata within the specified time limit you can demonstrate that you have adhered to the plan. It is also sensible to generate metadata as you go along, when it is fresh in everyone's mind, rather than wait until the end of the project. In a similar way you can check that backups have been carried out by their presence or absence.

Funding bodies may also wish to check that data management plans have been followed by way of written statements or other such methods for assuring adherence. Final deposit into a national archive, such as The UK Data Archive, also demonstrates adherence.

Adherence to this data management plan will be demonstrated by checking that tasks outlined have been carried out no later than two weeks after a sub-study, or set of experiments, has ended. For example, descriptive and other metadata should be constructed within two weeks. Backups will be checked on an ongoing basis to ensure they have taken place and are recoverable as part of a data recovery plan. At the end of the project deposit into The UK Data Archive will demonstrate adherence to planned exit strategy and long term preservation. Short internal reviews will be carried out on a monthly basis.

8.1.2. Who will do this?

For larger projects there may a project manager who is able to check data management plans at regular intervals to demonstrate that the statements outlined have been, and continue to be, met. For smaller projects this duty is likely to fall to the Principal Investigator. Some departments and host institutions my have a data specialist who can help verify that the data management plan is being adhered to. It is also possible that peer review by the funding body themselves could take place.

The Principal Investigator (J Ridley Stroop).

8.2. Review

8.2.1. When will this data management plan be reviewed?

For smaller, or shorter duration, projects it is likely that plans may only be reviewed once or twice. For larger long duration projects, or for those that are more complex, reviews may occur more frequently. There are no hard and fast rules on how frequently reviews should take place. Six monthly reviews probably offer a balanced approach. For shorter projects these may need to take place more frequently.

Data management plans will be reviewed on a rolling six month schedule.

8.2.2. Who will do this?

For larger projects there may be a project manager who is able to review data management plans at regular intervals to ensure they remain reliable and valid. For smaller projects this duty is likely to fall to the Principal Investigator. Some departments and host institutions my have a data specialist who can help review the data management plan. It is also possible that peer review by the funding body themselves could take place.

The Principal Investigator (J Ridley Stroop).

8.2.3. Does this version supersede an earlier plan?

If this plan supersedes an earlier one answer "Yes". If it does not, answer "No". You may also wish to highlight the title, version, date and author of any plans that are superseded for clarity and archival purposes.

No.

9. Agreement ratification by stakeholders

9.1. Statement of agreement (with signatures if required)

In much the same way as consortium agreements are used for multi-site multi-partner projects to articulate who is responsible for what aspects of a project, a statement of agreement to adhere to a data management plan is also a good idea. It is possible that a short statement can be included within a standard consortium agreement. It might merely state that all parties agree to carry out data management in accordance with the projects data management plan.

All parties agree to carry out data management in accordance with the project's data management plan.

10. Annexes

10.1. Contact details and expertise of nominated data managers/named individuals

In this section you should list any named people who have been allocated tasks in the data management plan. The named data manager might be the IT Manager within the department or host institution that is responsible for backups. Named individuals might be the Principal Investigator or Research Assistant. In terms of expertise it may suffice to list their job title and/or qualifications.

IT Manager, Department of Psychology, University of York, York, UK (Data Backup, Disaster Recovery & Network Provision)

J. Ridley Stroop, Principal Investigator, Department of Psychology, University of York, York, UK (Contextualising Documentation & Metadata Creation)

A Research Assistant, Department of Psychology, University of York, York, UK (Data Transformations)

10.2. Glossary of terms

If you have used any specialist terms or acronyms you are encouraged to include them in this section.

E-Prime: Software used to construct psychology experiments, present stimuli and collect data from participants

E-DataAid: Software used to analyse data collected by E-Prime

10.3. Other annexes as required

Rather than include large sections of text within the data management plan you are encouraged to append them as annexes. For example, annexes can include standard policies that are already in place at a host institution regarding data backup and recovery.