

# **Data Management Plan**

**example:**

**Economic and Social Research  
Council**



This DMP, made public with the kind permission of the PI Andrea Holomotz, represents a real example of a funded proposal from the University of Leeds that forms part of a wider suite of documentation submitted to the research funder. The contents of this DMP should not be directly replicated as a DMP is a document unique to your research proposal. Please take advice from your Faculty IT Manager and/or the Research Data Leeds Team (researchdataenquiries@leeds.ac.uk) on your requirements prior to making a DMP submission to your research funder.

## **Realist Evaluation of Adapted Sex Offender Treatment Programs for Men with Intellectual Disability**

Project Stage: Basic

RCUK Research Councils: Economic and Social Research Council

Lead organisation: University of Leeds

### **1 An explanation of the existing data sources that will be used by the research project (with references).**

There is some research that evaluates Adapted Sex Offender Treatment Programs (ASOTPs), but not necessarily by drawing on staff and especially user perspectives. A review of existing datasets, including searches on the UK Data Services website, found no published datasets on sex offenders with intellectual disability (ID) and/ or treatment programs.

The British Crime Survey can contextualise the research to some extent by exposing the prevalence of sexually violent crimes, but this is not disaggregated by whether or not the offender had an ID.

### **2 An analysis of the gaps identified between the currently available and required data for the research.**

ASOTPs have not previously been examined in the manner proposed by this research and there are no existing datasets in the archives which are adequate for answering the proposed research questions. There are no available datasets, which evaluate ASOTPs.

## **3 Information on the data that will be produced by the research project**

### **3.1 Data volume and data type, e.g. qualitative or quantitative data**

Qualitative data will be generated in audio format from interviews with practitioners (n=24 (phase 1&3) + up to 20 (phase 2)). There will be interviews (up to 20) and focus groups (4) with sex offenders with ID, which will be fully transcribed and anonymised. Fieldnotes of focus groups and also any drawing created by respondents with ID in focus groups and interviews will also be kept. This data will be anonymised and then processed and analysed using NVivo9. Quantitative data will be extracted from approximately 80 patient files and each case will be anonymised and then processed and analysed using SPSS.

### **3.2 Data quality, formats, standards documentation and metadata**

Audio files will be stored in MP3 or WAV format. Digital images will be stored as JPEGs. (Note that neither of these will be made available for data sharing, see section 7.) Microsoft Word 2007/2010 will be used for text based documents. .sav will be used for SPSS files.

These file formats have been chosen because they are accepted standards and in widespread use. At the end of the project, the Word documents will be converted to both plain text and PDF-A and long term preservation of the data from statistical analysis packages such as Stata will be carried out in accordance with the advice from the Council of European Social Science Data Archives ([http://www.cessda.org/project/doc/D10.4\\_Data\\_Formats.pdf](http://www.cessda.org/project/doc/D10.4_Data_Formats.pdf) section 4.3 pp33).

I am committed to providing high standard quality data and research excellence. To ensure the integrity and quality of the research data and increase the potential for data sharing, the transcriptions of the audio files will be checked and anonymised to make them ready for archiving. The formatting of data and the provision of metadata will conform to the UKDA standards and guidelines. This will also include clear data description, annotation, contextual information and documentation, e.g. unique identifier for each transcript, uniform and consistent

layout throughout data collection, cover sheet with interview details such as date, place and interviewee details. (Note that details will be kept vague to preserve anonymity, see section 7.)

### **3.3 Methodologies for data collection**

Semi-structured interviews and focus groups will be digitally recorded and subsequently transcribed into Microsoft Office Word 2010/2007. Fieldnotes on focus groups will be kept in Microsoft Office Word 2010/2007.

Quantitative data will be extracted from patient files and collated into SPSS.

A consistent system of file naming and an organised folder structure will ensure easy retrieval. This will involve creating meaningful but brief names and using file names to classify types of files.

## **4 Planned quality assurance and back-up procedures (security/storage)**

Electronic data will be stored on the University of Leeds SAN (Storage Area Network), which comprises enterprise level file servers in physically secure data centres with appropriate fire suppression equipment. Snapshots are taken every day at 10pm (and accessible for 1 month). A second level of snapshots is taken every month and are kept for 11 months. Snapshots are user recoverable from the desktop.

An incremental copy to backup tape is taken every night (and kept for 28 days) and a full copy is taken every month. Every quarter, the full dump tapes are moved to a long term storage facility where they are kept for 12 months.

Tapes are initially stored in on-campus fireproof safes and then moved to off-campus secure locations.

Access to electronic data is controlled by Active Directory (AD) Group membership. The Faculty IT Manager will set up a dedicated folder for this research project and create read-only and read-write AD groups. I will decide which users require read-only and read-write access. Off-campus access is via the Citrix portal.

External users who need access to the data will apply for a University username and then be assigned to the appropriate AD group.

Any sensitive data (as defined by the Data Protection Act) that is stored on portable electronic devices will be protected by encryption software to FIPS 140-2 standard. Any sensitive data that needs to be transmitted electronically will first be encrypted to FIPS 140-2.

If any highly sensitive data needs to be stored, then a research data folder on the SAN will be encrypted, so it can only be accessed by authorised members of the project with the appropriate encryption software installed on their desktop PCs. Highly sensitive data is not available from off-campus.

## **5 Plans for management and archiving of collected data**

As required by ESRC, this data management plan seeks to prepare the project data for future sharing and potential secondary analysis. This is particularly important in this project, as sex offenders with ID have thus far often been the objects and rarely the subjects of research. Therefore, the data (including anonymised transcripts of interviews where I have explicit permission for these to be used, but not including any audio or visual files) will be deposited for archiving and re-use with the ESRC data service provider, UKDA, at the end of the project and within three months of the end of the award. The data management plan will be reviewed during the life of the project to ensure the success of the long-term strategy. Prior to archiving, the data files will be converted to suitable open formats for long term preservation as described in section 3.2.

## **6 Expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties.**

As much of the data is generated through interviews with human participants, the ability to make it available for reuse will be subject to receiving the necessary level of consent from the individuals involved. Due to the highly sensitive nature of this research it is essential that the identities of participants remain concealed. However, ethically it is preferable to re-use data on

vulnerable populations, such as sex offenders with ID, rather than having to re-visit these for each individual research project. I will therefore keep a close eye on the data that is being produced during this research and seek to find ways of making it available for other researchers. One way could be to hand-pick a number of sample transcripts in partnership with respondents and edit these further for data sharing.

I have consulted with UK Data Service on possible strategies as part of my early data management planning. Appropriate restrictions to data sharing offered by the UK data archive will be put in place. I will furthermore attend a "Managing Sensitive Data in an Open Access Age" workshop and any other relevant training run by the UK data archive to find out more.

## **7 Explicit mention of consent, confidentiality, anonymisation and other ethical considerations**

Fully anonymised data arising from interviews and focus groups will not be shared, unless explicit consent was given by respondents. As sex offenders with intellectual disability (ID) are such a vulnerable group, particular care will be taken. Only transcripts that were especially prepared for this will be shared. As discussed in section 6, one way of handling this sensitive data could be to hand-pick a number of sample transcripts in partnership with respondents and edit these further, to remove any identifies, such as very peculiar phrases used by a person that makes them easy to identify or in fact mis-identify or lengthy descriptions of events or personal stories that may give a person away. Focus group data may not be shared, unless all participants consented to data sharing. However, editing may be applied to remove respondents who did not consent to data sharing from a focus group transcript, to allow at least some of the data to be shared.

Enabling people with ID to give their informed consent to something as complex as data sharing poses considerable challenges. I have experience of working with this group professionally and have also conducted past research with this group. I am able to break down complex concepts to this population and will put considerable effort into developing an accessible way to explain "data sharing". Consent for data sharing will be sought from the respondent in the presence of their key worker, who will sign the consent form as a witness. If they are in doubt about the respondent having understood "data sharing", the key worker will make a decision about this on their behalf, unless the respondent opted out. In that case their choice will be respected, even if they do not appear to have understood "data sharing" fully.

NHS and "Kantonale Ethikkommission" (Switzerland) approval will be sought regarding the sharing of fully anonymised and quantified data extracted from patient files in .sav format.

## **8 Copyright and intellectual property ownership of the data**

The intellectual property of the data generated will remain with the University of Leeds. However, the University policy of the management of research data requires all data arising from research projects to be made openly available where possible. The research will not use any data which is covered by the Copyright, Designs and Patents Act 1988 or any other similar legislation.

## **9 Responsibilities for data management and curation within research teams at all participating institutions**

I will have overall responsibility for implementing the data management plan. The Faculty IT Manager will be responsible for ensuring that electronic file permissions have been correctly assigned and for advising on other aspects of data storage and security. Staff involved in the project at participating organisations will be responsible for following data management procedures. The data management plan will be monitored in meetings with my experienced mentor and during RA supervision.