

APPENDIX O: The Big Picture: Data Management Plan

(Aspects related to data sharing are highlighted below)

1.0 Assessment of existing data

The UK Data Service offers large scale UK survey data on adolescent substance use, and some data on Indian mental health, but only in relation to gender, violence and poverty. There is no data on the lived experience of young Indian people at risk of, or recovering from, mental health difficulties linked to substance use. We have, however, drawn on the *National Mental Health Survey of India, 2015-16* to inform our project plans and will return to this to situate our project findings.

2.0 Description of project data

Data will be generated via: Photovoice (PV) with young people; stakeholder interviews and mapping; fieldwork on capacity building; and project evaluation. From these, the project will generate **five forms** of data.

(1) PV **images** will be either *participant-generated* images (anticipated to be photographs taken on smartphones or disposable cameras; RF will supply Kodak 39 exposure single-use cameras) and/or *participant-collated* images (e.g., from personal archives, the internet, magazines). We estimate a project total of 250-300 images (i.e. ~10 per PV participant).

(2) **Audio and transcribed interview data** from PV (n~30) and interview data from policy makers (n~15) in either English or Assamese. Audio interview data will be recorded on high quality, encrypted digital recorders. Transcription of interviews conducted in English will be via a professional service compliant with standard protocols (i.e., non-disclosure, transfer of encrypted files only, denaturalised transcript - verbatim 'play-script'). Transcription of interviews conducted in Assamese will be by the RF and will retain the original language. The RF will provide a summary of each Assamese interview in English.

(3) Policy makers' **database** will be manually generated by Col Mirzoev and PhD student.

(4) **Field records** on capacity building will be manually generated by the RF, PI CoIs & UoL-funded PhD student according to an agreed ESRC compliant protocol.

(5) **Evaluation** data will be collected by the Research Team and MIND India and will include questionnaires, website monitoring, social media activity, event attendees, 'vox pop' attendee interviews and new partnership logs. Data volume is difficult to specify in advance.

3.0 Data management, quality, documentation and curation

3.1 File Formats

We adhere to UK Data Service ESRC guidance on file formats.

(1) **Images:** Disposable camera images will be printed with copies for the participants and project. Participants' *collated* non-digital images will be either photographed (via the RF's smartphone) or scanned by the RF before interview end, and uploaded for project storage by the RF. These, along with *participant-generated* digital images, will be stored as TIFF or JPEG depending on curation and reproduction needs. The RF will assign unique identifiers and time stamps to mark when in the interview each image was discussed / referenced.

(2) **Audio and transcribed interview data** will be deleted from the recorder following safe upload and storage of the master files. These will be assigned unique identifiers by the RF and stored as Waveform Audio Format files or Free LossLess Audio Codec. Transcription will be in MS Word files.

(3) Policy makers' **database** will be an MS excel file.

(4) **Field records** on capacity building activities will be MS Word files.

(5) **Evaluation data** will be stored as MS word or excel files.

At **project end**, for long term preservation, Word documents will be converted to plain text and PDF-A and MS excel files will be exported, depending on data complexity, to either plain text, tab separated or excel tables by the Project Administrator with PI guidance.

3.2 Anonymisation procedures

Participant **personal data** will be anonymised by the Research Team member generating the data. Data will be assigned a unique identifier, created by the participant according to given (retrievable) instructions, known only to them and the Research Team. The only exception to this is the policy mapping database as this information (i.e., roles, institutions and work contact details) is in the public domain. Interview **transcripts** will have direct and indirect identifiers (age, community, family, etc.) and any other sensitive material removed prior to storage. Images will be stored for analysis in original form. However, prior to **images** entering public domain (if consent is given), identifying features of people, other identifying markers (e.g. rehabilitation centres), and any potentially sensitive material will be removed.

3.2 Quality assurance

With regards to the qualitative data, the RF, PhD student, and, where appropriate, the Project Administrator will: (i) check the accuracy of transcription against audio; (ii) ensure document coversheets for each participant have correct identifiers and consent record for data to enter the public domain; (iii) ensure anonymization and obscuring, cross-checked by PI or Col Hugh-Jones; (iv) assign PV image to a participant code and to a time point in the transcript where it was discussed, cross-checked by Administrator; (v) with the PI and Col Hugh-Jones, follow published guidelines for thematic analysis (Braun & Clarke, 2006) and framework analysis (Ritchie & Spencer, 1994) as required (see *Case for Support*); (vi) complete audit tables; and (vii) with Col Hugh-Jones, use quality evaluation frameworks for interview research (Leonikdaki, 2015) and qualitative analysis (Tracy, 2010). Database entry mapping policy makers will be cross-checked by PhD student and Administrator and checked for completeness by Col Mirzoev. Fieldnotes on capacity building will be checked for completeness by PI. Evaluation data entry will be by Administrator, cross checked by RF.

3.3 Data documentation standards

Each dataset will be accompanied by contextual descriptions and metadata, meeting UK Data Service standards and FAIR principles, operationalised in a protocol established by Research Team early in the project. PI will have oversight, Research Team members responsible for following protocol wrt the data they generated, support by Administrator. For qualitative data, we will follow standardised protocols and report **substantive and methodological context** data were generated in to support appropriate interpretation, although it is acknowledged that full complexity cannot be objectively archived. Contextual descriptions and metadata will include: project aims, objectives, investigators and funders; work strands; structure of data files; descriptions of data provenance; methods used for data generation (including recruitment processes, ethical safeguards and interview schedules); data preparation procedures (including transcription, checking, anonymising, obscuring and/or cleaning); descriptions important to appropriate data interpretation; image tagging with description of images given by participants. Interviews conducted in Assamese will be accompanied by an English summary created by RF. PI will document publications, presentations and other research outputs that explain or draw on the data.

3.4 Data Storage

The Research Team will store low sensitivity data (i.e. non-personal, low risk) on the approved University of Leeds (UoL) OneDrive. Only the UK Research Team and the RF in India will have authorised access. High sensitivity data (i.e. the PV images, and all interview audios and transcripts) will only be stored on UoL servers (backed up daily) in password-protected files, accessible by the RF, Administrator, PI and Col Hugh-Jones. The RF has access to the UoL server from India via Citrix portal and has already successfully uploaded audios and images to there from Assam. During transcription of Assamese interviews by the RF, she will store the files as encrypted and password protected on a project laptop, before being transferred to UoL server and then deleted from the local device. Should the need arise, sensitive data can be shared through OneDrive using encryption. For all data storage, we will use a consistent system of file naming and an organised folder structure. The Research Team will establish a protocol for **version control and edit histories**, and the PI will have responsibility for **master file** protection and master file destruction at the designated time. The PI will allocate edit or read-only permissions. We will not store data overseas and data will not be sent overseas via email. However, email will be used within India to allow participants to send images to the RF for PV interviews. We will apply the UoL's 'Risk Assessment for Security of Data' at key points throughout the project.

4.0 Data archiving and sharing

The PI will store project evaluation data for 3 years and remainder of data on UoL OneDrive and UoL servers for 10 yrs. We have planned data sharing from outset of project planning.

4.1 Permission for sharing

Most of the data will be suitable for sharing given our consent procedures, our data anonymization procedures and the detail of our contextual and metadata. Our planned consenting procedure will explain to participants that they will have the **option to exclude** their anonymised data (transcripts and images) entering the public domain, i.e. on our Big Picture website, on ReShare or in published reports. They will be able to opt in or out of each of these distinct options, and both for images (on a case by case basis) and narrative separately. This is an important offer given the likely perceived inequalities in power between the researcher and the researched, as well as likely fears about being investigated and documented. Participants will be asked to make a final decision on data sharing by RF at interview end (considering their disclosures). RF will record written consent, or audio consent where written consent is feared or impossible. We will not share the PV images or interview data without participant consent and / or where we believe anonymisation does not fully protect the individual or poses a risk to others.

4.2 Timelines and accessibility

PI will deposit the data for archiving and re-use with UK Data Service ReShare and within three months of the end of the award, with a 12 month embargo for public access to permit the Research Team to publish. Prior to archiving, data files will be converted by Administrator to suitable open formats for long term preservation as described above. Non-sensitive information such as the field notes and policy mapping will be **open-data**. It is likely that the PV data will also be open-data but PI will review with our Steering Group and the UK Data Service whether a **safeguarded** position is more suitable once the full extent of the data is known, and its accordance with the ICO's Code of Practice on Anonymisation. If a safeguarded option is adopted, we will establish a data sharing agreement for sensitive elements of the data (aligned with our approved ethics protocol and participant consent). PI will publish our data sharing policy on our project website, along with the location of the data in ReShare at project end. This location will also be reported in publications arising from the project. PI will ensure accurate project reporting on RCUK Gateway to Research. The Research Team will test the discoverability of the data to ensure effective metadata.

5.0 Copyright and Responsibilities

The intellectual property of the data generated will remain with the UoL. Study-wide data management will adhere to the UoL Data Management Policy & Procedures (<https://library.leeds.ac.uk/research-data-policies>) and Data Security Policy (http://it.leeds.ac.uk/info/116/policies/249/information_protection_policy). Study-wide data management will be the responsibility of the Research Team, supported by UoL Research Data Leeds Service. Metadata will be created by PI and Col Hugh-Jones. Data quality assurance will be managed by the Research Team with ultimate responsibility held by the PI. The data management plan will be reviewed during the life of the project to ensure the success of the long-term strategy.

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Leonidaki, V. (2015). Critical appraisal in the context of integrations of qualitative evidence in applied psychology: The introduction of a new appraisal tool for interview studies. *Qualitative Research in Psychology*, 12, 435-452.
- Ritchie, J. & Spencer, L. (1994). Qualitative data analysis for applied policy research (pp. 173-194). In Bryman, A. Burgess, R. (Eds), *Analyzing Qualitative Data*. NY: Routledge.
- Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, 16, 837-851.