Aaron Tillekeratne

taru3@student.monash.edu

Abstract

This document contains the data management checklist for the MP4D research projected aimed at develop current, accurate, and trustworthy interactive information (primarily via mobile technologies) for women in agriculture in rural areas, to enable them to act to improve their well-being and livelihoods in one of the poorest and ecologically fragile parts of the world.

MP4D

Data Management Plan Checklist

# Stakeholders

|  |  |
| --- | --- |
| Stakeholder | Role |
| AusUni | Research team |
| NGO | Facilitator and project owner |
| Data Security Contractor | Secure data storage for project data. |
| Media Outreach Company | Mobile phone based data capture and provider including aggregation and deidentification service. |
| Participant Villages (2) | Participants in research project. |
| Control Villages (2) | Control group villages in research project. |

**Research Data Planning Checklist**

|  |  |
| --- | --- |
| **Name** | Aruna (Aaron) Tillekeratne |
| **Faculty / division / unit** | Data Science |
| **Contact details** |  |
| **Title / description of the Graduate Research project**  MP4D | |

|  |  |  |
| --- | --- | --- |
|  |  | **A. DATA** |
|  |  | **Please complete the following table listing all data files to be included in this DMP. Include data produced by the project, data acquired from third party sources, and data required for understanding the context or conduct of the project. Please identify any publications that may result from the project, and indicate whether or not they are to be included with the collection. Add additional rows as required.** |
| 1. |  | |  |  |  |  | | --- | --- | --- | --- | | **Data File** | **Format** | **To be archived (Y/N)?** | **Open access (Y/N)?** | | **Data Files Produced by the project.**  Project Scope Documents | Electronic (DOCX) | Y | N | | Project Schedule | Electronic (proprietary format) | Y | N | | Survey Form Template | Electronic (DOCX) | Y | Y | | Completed Survey Forms | Paper, Electronic (DOCX) | Paper – N, Electronic – Y | N | | Ethics & Consent Form Template | Electronic (DOCX) | Y | Y | | Completed Ethics and Survey Forms | Paper, Electronic (DOCX) | Paper – N, Electronic – Y | N | | Mobile Phone Voice Data | Electronic (.WAV) | Y | N | | Mobile Phone Text Data | Electronic (CSV) | Y | N | | Mobile Application Data | Electronic (CSV) | Y | N | | Mobile Photo Data | Electronic (TIFF) | Y | N | | Training Session Templates | Electronic (DOCX) | Y | Y | | Training Session Participant Forms | Paper, Electronic (DOCX) | Paper – N, Electronic – Y | N | | Training Session Photos | Electronic (TIFF) | Y | N | | Group Meeting Agendas | Electronic (DOCX) | Y | Y | | Group Meeting Notes | Paper, Electronic (DOCX) | Paper – N, Electronic – Y | N | | Group Meeting Photos | Electronic (TIFF) | Y | N | | Workshop Template | Electronic (DOCX) | Y | Y | | Workshop Records | Electronic (DOCX) | Y | N | | Workshop Photos | Electronic (TIFF) | Y | N | | Interview Notes | Paper, Electronic (DOCX) | Paper – N, Electronic – Y | N | | Interview Audio Records | Electronic (.WAV) | Y | N | | Mobile Phone Voice Aggregated Data | Electronic (CSV) | Y | Y | | Mobile Phone Message Aggregated Data | Electronic (CSV) | Y | Y | | Mobile Phone Aggregated Application Data | Electronic (CSV) | Y | Y | | Mobile Phone Photo Aggregated Data | Electronic (CSV) | Y | Y | | Analytics Scripts | Electronic (TXT) | Y | Y | | Models | Electronic (XML) | Y | Y | | Analytics Tabular Data | Electronic (CSV) | Y | Y | | Analytic Reports | Electronic (DOCX) | Y | Y | | **Publications**  Dataset | Electronic (CSV) | Y | Y | | Journal Article | Electronic (DOCX) | Y | Y | | Report to NGO | Electronic (DOCX) | Y | N | | Book Chapter | Electronic (DOCX) | Y | Y | | Report to Participants | Electronic (DOCX) | Y | Y | | Conference Report | Electronic (DOCX) | Y | Y | |
|  |  | **A. OWNERSHIP, COPYRIGHT, INTELLECTUAL PROPERTY (IP)** |
|  |  |  |
|  |  | **Copyright protection** |
| 2. |  | The data is protected by copyright.  *This will apply to most research data.*  As per Copyright Act 1968, the outputs will be subject to copyright protection. |
| 3. |  | The data will be collected, created or compiled |
|  |  | * in Australia - Australian copyright applies.   All research analysis outputs and findings. |
|  |  | * outside of Australia.   Data collection and deidentification. |
|  |  | **Ownership of copyright and IP** |
| 4. |  | The copyright and other IP in the data is owned by: Jointly owned by NGO and AusUni. Copyright level may change through project progression. |
|  |  | Monash University  I have assigned IP to the University because it falls into one of the categories prescribed under the Statutes and Regulations. |
|  |  | Monash University (joint ownership)  Research conducted by Monash in collaboration: copyright and IP ownership are documented in an agreement between the organisations.  Reference:Refer NGO Project Scope document for full details. |
|  |  | Someone else owns the data |
|  |  | □ Monash Graduate Research student is creating/collecting the data, but another party owns the copyright and IP.  Reference: |
|  |  | □ Monash Graduate Research student is using data originating from another party and that party owns the data.  Reference: |
|  |  |  |
|  |  | **Third party data** |
| 5. |  | If someone else owns the data, how did you obtain it and what terms and conditions apply to your use of it?  All data used in the research is gathered within the project in partnership with NGO and AusUni. |
|  |  | Purchased or licensed the data commercially  Reference: |
|  |  | Obtained data under an open access license  Reference: |
|  |  | Obtained data through other means  Details: |
|  |  |  |
|  |  | **B. ETHICAL REQUIREMENTS** |
|  |  |  |
| 6. |  | The research involves human subjects.  This includes re-use of data from or about people, e.g. from health agencies.  The research conduction involves human subjects and any data collection, directly, observed or otherwise will only be recorded after the submission of informed consent form. |
| 7. | □ | A Human Ethics Application has covered / will cover requirements in the following areas: |
|  |  | * Privacy   + No collected data will be made available to the research team, NGO team, public or any other stakeholder without deidentification. * Confidentiality   + Any data that does not violate the privacy requirement shall be constrained by confidentiality. All data, information or outputs from the project is to be shared wholly by the research team. Non-project stakeholders will not have access to this information. * Cultural sensitivity   + Culturally sensitive data is not expected to be gathered or any cultural inferences obtained in the analysis process is to be thoroughly evaluated by the research team before publication. * Other   + No other special ethics concerns apply.   Ethics application ref. no.: |
| 8. |  | Other special requirements for managing data have been negotiated in addition to the Human Ethics Application.  Details:  NGO contractors are responsible for providing security and redundancy for all recorded data. All persistent data is deidentified in order to maintain privacy of the participants. |
|  |  |  |
|  |  | **C. DURABLE FORMATS** |
|  |  |  |
| 9. |  | Data formats used:   |  |  | | --- | --- | | **Data Type** | **Format** | | Documents & Text Documents | DOCX | | Tabular Data | CSV | | Image Data | TIFF | | Audio Data | WAV | | Application Data | RDBMS\* |   Details: \* Relational Data Store is used in the project to facilitate access data captured from mobile phones. The data goes through deidentification from the Media Outreach Company (contracted by NGO) to provide mobile phone data from participants. |
| 10. |  | Data formats meet the following criteria: |
|  |  | * Endorsed by an international or national standards agency * Widely used * Accepted as best practice in this discipline   Details: All data formats are durable, non-proprietary and not platform dependent. |
| 11. |  | Special hardware or software requirements (for digital data)  Details: Provided by research partner NGO. |
|  |  |  |
|  |  | **D. STORAGE AND BACKUP** |
|  |  |  |
|  |  | **Digital data** |
| 12. |  | Digital data is stored in: |
|  |  | Large Research Data Store (LaRDS)  Location: |
|  |  | Faculty-allocated network storage (e.g. “V-drive”)  Location: |
|  |  | Other  Details: Secure file and relational database system provided by the NGO. |
| 13. |  | Digital data is backed up: |
|  |  | On LaRDS – automatic nightly backup |
|  |  | On faculty-allocated network storage  Frequency of backup: |
|  |  | Other backup arrangements  Details: Automatic server backup by the NGO’s data services contractor including geo-replication. |
|  |  |  |
|  |  | **Non-digital data** |
| 14. |  | Data in non-digital formats is/will be stored in: |
|  |  | Secure facilities located in the school, institute, or centre  Location: |
|  |  | Other  Location: Secured at NGO headquarters, only for duration of the research component. All physical data to be digitised and then destroyed. |
|  |  |  |
|  |  | **Data volumes (all formats)** |
| 15. |  | Estimated data volumes: < 500GB  Details: All data, including images and audio is not expected to exceed 500GB. |
|  |  |  |
|  |  | **E. SHARING DATA AND CONTROLLING ACCESS** |
|  |  | *This section relates to data sharing during the course of the project. Longer-term data sharing is covered in the section below on deposit in a repository or archive.* |
|  |  |  |
|  |  | **Sharing data during the project** |
| 16. |  | During the project, data will be shared |
|  |  | Not at all – I am the only person that will have access |
|  |  | Internally – Monash staff (e.g. supervisor) and other students |
|  |  | Externally – e.g. co-supervisor at another institution, research participants, funding agency)  Details: The AusUni research team and NGO staff (authority controlled by NGO) will be the only persons allowed to access during the research process. |
| 17. |  | Data will be shared with external parties in the following ways: |
|  |  | Data stored at AusUni will be accessed by others, e.g. using an online shared workspace  Details: |
|  |  | Data will be transferred to the other location/s |
|  |  | * using specialised data transfer tools like file transfer protocol (FTP) , or GridFTP   + All electronic documentation is to be transferred from gathering location to NGO file servers via SFTP. * by email   + Absolutely no data to be transferred via email. * on physical media, e.g. DVDs, hard disk drives   + Absolutely no data is to be transferred on any physical media. * by other means   Details: All physical documentation is to be mailed using the NGO provided mailing service to NGO HQ. |
|  |  |  |
|  |  | **Controlling access** |
| 18. |  | Access to the data during the project will be |
|  |  | restricted |
|  |  | * password-protected (digital data only)   + Project folder will have a master password only revealed to the project team. * encrypted (digital data only)   + Mobile phone data (deidentified) is to be stored encrypted in the NGO provided relational data store. * physical security, e.g. locked filing cabinets, check-out procedures   + NGO HQ to provide dedicated project area with swipe card access to project team members. Documentation cabinets are not required to have individual locks.   Details: |
|  |  | unrestricted. |
|  |  |  |
|  |  | **F. DOCUMENTATION AND METADATA** |
|  |  |  |
| 19. |  | The following documentation and/or metadata (information about the data) will ensure data can be retrieved and used: |
|  |  | Inventory of data assets  Details:   |  |  | | --- | --- | | **Data** **File** | **Metadata** **Type** | | Physical files (e.g. survey forms and workshop forms, etc.) | Descriptive – NGO employee ID numbers must be attached although removed from public viewing. | | Images | Descriptive, Technical – A clear description of the event, subject as well as image dimensions must be available. | | Phone Conversations | Technical – Standard metadata to be attached (see standard below). | | Text Messages | Technical – Standard metadata to be attached (see standard below). | | Mobile phone images | Technical – No data to identify the users, but standard image metadata only to be kept (see standard below). | | Aggregated and Analytic outputs | Descriptive – Full description of algorithm(s) or manipulations to be provided. | |
|  |  | Metadata standards  Details: AGLS Metadata Standard is to be used when designing individual fields. |
|  |  | Protocols for identifiers / reference numbers  Details: When digitising documents, a standard set of identifiers will be used, which include employee ID number of NGO curator who is responsible for the digitisation. Identifier for the stakeholder is also attached when data is provided to the NGO file server or the RDBMS server. |
|  |  | Protocols for filenaming  Details: A standard filename generator software application will be provided, where the data curator responsible for digitising or moving to SFTP server will be asked to login as well as provide some basic information about the data (data quality, type of document, etc.). Based on information a filename will be generated for each file with a clear dictionary of how the filename system made available to the whole project team. |
|  |  | Protocols for document structures – e.g. column headings, document properties etc.  Details: All document properties to include the source or the organisation and dates. |
|  |  | Data dictionaries, data definition files and schema  Details: RDBMS schema is to be provided by NGO of the relational data store, all file system dictionary is provided by the filename system designers and kept at the highest level for easy access. |
|  |  | Other documentation or metadata requirements (including software that may be used to managed documentation and metadata)  Details: Custom built filename generator with how the filenames are generated. Simple logic based program. |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  | **G. RETENTION AND DISPOSAL** |
|  |  |  |
|  |  | **Minimum periods** |
| 20. |  | Data must be retained after submission of the thesis or publication of results for a minimum of: |
|  |  | 5 years  Standard retention period |
|  |  | 7 years  Psychological testing or intervention with adults |
|  |  | 15 years  Medical research involving clinical trials |
|  |  | 25 years after date of birth of participants  Psychological testing or intervention with children |
|  |  | Other time period  Details: All data (deidentified) is to be kept for the duration of the project, post publication the raw mobile phone data will be disposed and only aggregated data will be retained. Any other data shall be retained for standard of 5 years of minimum after the completion of the project. |
| 21. |  | Based on the above, data must be kept until at least:  Date: 05/02/2023 |
|  |  |  |
|  |  | **Long-term and permanent retention** |
|  |  | *Can be completed by the researcher in draft form at the start of the project and should be reviewed regularly as research progresses.* |
| 22. |  | Data should be considered for permanent retention / archiving because the research |
|  |  | * is controversial   + Data/outcomes will not contain any controversial topics. * is of wide public interest   + Is of wide public interest in facilitating increased efficiency in agricultural reward for individuals. * uses an innovative technique for the first time   + Is a newly developed technology. * shifts the paradigm in this field of inquiry   + No significant change is expected in terms of the science approaching the problem. * would be costly or impossible to reproduce   + Scale of depth of the experiment would be difficult to reproduce as it involves experiments in remote areas with significant participation. * will be of enduring value to researchers in this discipline   + The data is valuable to researchers in the agricultural as well as technological industries. * will be of enduring value to researchers in other disciplines   + Significant value could not be identified to researchers of other fields. * supports a patent application or other formal IP process   + It is the intention of the NGO to develop some product using the outcomes of the research project. * Other   + No other special conditions apply.   Details: |
| 23. |  | The ability to permanently retain / archive the data and make it accessible will be affected by: |
|  |  | * legal issues around ownership of copyright and IP   + There may be some copyright issues with the outcomes of the research as stipulated by the NGO and its intent to use in a product or service in the uture. * ethical requirements for data to be destroyed   + No ethical requirement of data destruction is identified. The destruction of data will not impede any researchers or the public. * legal or ethical requirements for access to data to be restricted   + Given that there is voice data present, it is important to restrict the data to the research team. Once destroyed it may no longer have any ethical or legal restrictions. * technical issues, e.g. obsolete data formats or software   + Durable formats are implemented and there is no current concern with obsolete data or format issues. * Other   + No other concerns around data retention at this time.   Details: |
|  |  |  |
|  |  | **H. DEPOSIT IN A REPOSITORY OR ARCHIVE** |
|  |  | Before completing the next section, please list the features of a data archive that would be suitable for this dataset   * Ability to contain large number of individual files of differing formats. * Ability to provide some sort of relational data environment if possible. |
| 24. |  | The data will be deposited in a repository or archive. |
|  |  | * AusUni University Research Repository   + Post publication, data will be deposited in the AusUni repository for access by the public. |
|  |  | * another repository or data archive   Details: |
| 25. |  | I have contacted the archive / repository manager and understand the terms and conditions of deposit. |
|  |  | In order to support discovery of your data, please create a draft RIF-CS with all the required metadata to  describe the data collection that will be archive: See … for details of the required metadata fields  Collection:   |  |  | | --- | --- | | **Label** | **Value(s)** | | Registry objects | MP4D Data | | Registry Object | File Data, Analytics Data | | Originating Source | MP4D Research Project | | Group | Research Input Data | | Key |  | | Class Type | Files, Tabular | | Name | Survey Data, Voice Analytics Data, Text Message Data, Image Data, Form Data |   Party:   |  |  | | --- | --- | | **Label** | **Value(s)** | | Registry objects | MP4D Parties | | Registry Object | Facilitators, Researchers, Participants, Control Groups | | Originating Source | MP4D Research Project | | Group | Stakeholders | | Key |  | | Class Type | Party | | Party Type | Facilitators, Researchers, Participants, Control Groups | | Name | NGO, AusUni, Group 1, Group 2, Group 3, Group 4, Group 5 |   Event:   |  |  | | --- | --- | | **Label** | **Value(s)** | | Registry objects | MP4D Data | | Registry Object | Survey, Interview, Workshop, Group Meeting, Training Session | | Originating Source | MP4D Research Project | | Group | Research Activity | | Key |  | | Class Type | Activity | | Activity type | Survey, Interview, Workshop, Group Meeting, Training Session | | Name / Title | Baseline survey, Post implementation survey, Control group training session, | | Description | Survey conducted prior to start of work, training session using mobile phones. |   Service:   |  |  | | --- | --- | | **Label** | **Value(s)** | | Registry objects | MP4D Data | | Registry Object | Analytics Data, Survey Data | | Originating Source | MP4D Research Project | | Group | Data Service | | Key |  | | Class Type | Service | | Service type | Data Services | | Name | File Service, RDBMS Service | |
| 26. |  | Data will **not** be deposited in a repository or archive. |
|  |  | * Data and documentation stored on LaRDS will be associated with the academic unit. |
|  |  | * Data and documentation stored on faculty drives will be associated with the academic unit. |
|  |  | * A copy of local data (e.g. from a personal laptop or home computer or on removable media) will be provided to the academic unit.   Details: |
|  |  | * I will keep copies of the data myself.   Details: |
|  |  | * Data will be securely destroyed.   Details: |
| 27. |  | I have discussed these arrangements for long-term care of the data with my academic unit and other relevant staff.  TBC |