RESPIRE Data Management Plan (DMP): Template (adapted from the University of Edinburgh)

Name:	Professor Dr Ee Ming Khoo
Project Title:	The Malaysian Asthma Hajj Study
Institute:	University of Malaya, Kuala Lumpur, Malaysia
Start Date:	1st August 2018
End Date:	31st January 2021
DMP version number and date:	DMP Version 1 and 01.12.2020

Responsibilities & Resources (applicable across the sections below)

Who will be involved in the data management of this research?

Person in charge:
Principal investigator, principal investigator at the site and research manager of the project are listed below:

1. Principal Investigator: Professor Dr Ee Ming Khoo
2. Principal Investigator at the site: Professor Dr Su May Liew, Dr Ahmad Ihsan Bin Abu Bakar
3. Research Manager: Jayakayatri Jeevajothi Nathan

Resources:
Hardware and software used in the management of data are listed below:

1. Laptop (encrypted and secured with username and password and kept in a locked cabinet)
2. Thumb drive (encrypted, password protected and kept in a locked cabinet)
3. Solid-state drive (encrypted, password protected and kept in a locked cabinet)
4. DataStore (to back up active research data)
5. Microsoft Word
6. Microsoft Excel
7. NVivo Qualitative Data Analysis Version 12.0 Software
8. SPSS Software version 25.0
9. Locked cabinet for any hard copies (research protocol, topic guide, participant information sheets, sociodemographic forms, informed consents, questionnaires, codebook and NVivo coding framework)

1. Data Capture

What data will be generated or reused in this research?

There are 3 phases in this project. Data that were or will be generated are listed below:

1. Phase 1: Prevalence of asthma among pilgrims, asthma control and incidence of unscheduled care and hospitalisation
   (a) For the whole hajj period
   The whole hajj period data will be retrieved in the form of manual/electronic records from the Pilgrim Board (still waiting for approval).
   (b) For the asthma sub-cohort
   Quantitative - sociodemographic and personal profiles, asthma status, other medical conditions, asthma treatment and management, clinical examination such as peak flow rate, unscheduled visit and hospitalisation status and asthma management.
2. Phase 2: Observation of the organisation of the pre hajj reviews
   Qualitative transcript of public and private practice.
3. Phase 3: Perceptions of the stakeholders on the risk of asthma during the hajj
   Qualitative transcript of healthcare provider, policy makers, manager and returning pilgrims from Hajj.

How much data will be generated?

50GB – 250GB

2. Data Management

How will the data be documented to ensure it can be understood?

1. Phase 1: Prevalence of asthma among pilgrims, asthma control and incidence of unscheduled care and hospitalisation
   (a) For the whole hajj period
   Source documents will be documented in Microsoft Excel and SPSS (still waiting for approval form Pilgrim Board to retrieve the record).
   (b) For the asthma sub-cohort
   Source documents were documented in Microsoft Excel and SPSS. A codebook containing information about each of the variable in the dataset were created to ensure that the data is understood and interpreted properly.
2. Phase 2: Observation of the organisation of the pre hajj reviews
   Source documents were transferred into Microsoft Word and exported to Microsoft Excel then into NVivo Qualitative Data Analysis Version 12.0 Software for analysis.
3. Phase 3: Perceptions of the stakeholders on the risk of asthma during the hajj
   Each interview was assigned with a participant code and transcribed verbatim in Microsoft Word. The de-identified interview transcripts were exported into NVivo Qualitative Data Analysis Version 12.0 Software for data analysis.

Where will the data be stored and backed-up?

The electronic data were stored in two RESPIRE laptops, a thumb drive, a solid-state drive and will be backed-up in DataStore. Hard copies (research protocol, topic guide, participant information sheets, sociodemographic forms, informed consents, questionnaires, codebook and NVivo coding framework) were stored in locked cabinet.

3. Integrity

How will you quality assure your data?

The completed questionnaires were checked by researcher for completeness of the data before it was entered into SPSS Software version 25.0. Data entered were checked for errors by running a frequency distribution on each of the variables. Participants would be contacted for clarification if any missing data or error traced.

For observation phase, at least two researchers were present at the study site and the notes were cross checked to ensure the observation were matched.

The interviews were recorded by using two audio recorders and the non-verbal cues were noted down by a researcher. Interviews were transcribed verbatim and the transcripts were checked by two researchers for accuracy.

4. Confidentiality

How will you manage any ethical and Intellectual Property Rights issues?

1. The study will be conducted in accordance with the principles of the International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice (ICH GCP). All required approvals have been obtained.
2. The investigator is responsible for the overall conduct of the study at the site and compliance with the protocol and any protocol amendments.
3. All reports and other records will be identified in a manner designed to maintain participant confidentiality. All records will be kept in a secure storage area with limited access. Clinical information will not be released without the written permission of the participant. The Investigator and study site staff involved with this study may not disclose or use for any purpose other than performance of the study, any data, record, or other unpublished, confidential information disclosed to those individuals for the purpose of the study. Prior written agreement from the sponsor or its designee must be obtained for the disclosure of any said confidential information to other parties.
4. All Investigators and study site staff involved with this study must comply with the requirements of the appropriate data protection legislation including the General Data Protection Regulation

and Data Protection Act 2018 with regard to the collection, storage, processing and disclosure of personal information and will uphold the Act's core principles.

5. Retention and Preservation

Which data do you plan to keep and for how long?

Identifiable Data
All sociodemographic forms and informed consents of the participants will be kept for five years. Audio recordings will be permanently deleted at the end of the research project.

Non-identifiable Data
De-identified electronic data will be preserved for long term.

How will the data be preserved?

The hard copies of these identifiable data will be kept in a locked cabinet and will be shredded and disposed in secure bins after five years. The non-identifiable data will be preserved in two RESPIRE laptops, a thumb drive, a solid-state drive and on DataVault.

6. Sharing and Publication

Which data will be shared and how?

The final de-identified analysed data, conference abstract presentations, journal articles and dissemination event will be shared in DataShare.

Are any restrictions on data sharing required?

Yes, prior to release for sharing, there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, for data archived in Edinburgh DataVault, we will make the data available to users only under a data-sharing agreement that provides for:

1. A commitment to using the data only for research purposes and not to identify any individual participant;
2. A commitment to destroying or returning the data after analyses are completed.