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

Name:	Dr Osman Mohammad Yusuf
Project Title:	Development of ELISA and Rapid Testing Kits for COVID 19 and their Commercialisation in Pakistan
Institute:	The Asthma & Allergy Institute, Pakistan.
Start Date:	15 August, 2021
End Date:	30 November, 2021
DMP version number and date:	V-1.0 August 2020

Responsibilities & Resources (applicable across the sections below)

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

Data will be maintained, after vetting by the Principal Investigator and by the Data Management Team, under Dr. Aimal Rextin for entering, maintenance, and security.

1. Data Capture

What data will be generated or reused in this research?

The laboratory data will be generated in the research labs of The Asthma & Allergy Institute, (AAIP) Pakistan and analysed by the PI, Co-PI and invited specialists. The electronic data will consist of documenting the different diagnostic reagents imported, with their technical details, specifications and expiry dates. A record will be kept of the use of these reagents, and their performance characteristics in the research laboratory.

The second main data set will involve the development of the diagnostic kits, and documentation of all physical, chemical and immunological parameters around each diagnostic procedure, and the outcomes, and whether success or failure. The third set of data will be generated when the diagnostic kits developed will the tested against the known samples received from the commercial (Metropole) laboratories and the results compared.

The data of the development of the diagnostic kits, and the outcomes will be shared on a regular basis with RESPIRE / University of Edinburgh (UoE). The results generated will be evaluated and analysed by the specialists in the research team and statistically evaluated by the Data team of The Asthma & Allergy Institute, Pakistan, and these analyses will be shared and discussed with RESPIRE.

No other electronic information will be disseminated as there is no benefit in doing so.

How much data will be generated?

We have a sample size of 300. We assume that the outcomes of various laboratory procedures for one participant takes 5 KB. Then the data generated will be

$$300 \times 5\ KB \approx 2MB$$

2. Data Management

The data will consist of a single .csv file where each row corresponds to a single participant. The first column of each row will contain the anonymised ID of a participant and the remaining columns will contain the outcome of the different laboratory procedures that were applied to the sample. A cell will contain NA if the corresponding procedure was not applied on the sample from the corresponding participant.

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

We will prepare a technical report for the secondary users to understand the collection procedure and processing of the generated data with a bibliographical citation for users to cite in future publications. It will also contain full details of all laboratory procedures that were applied on the samples.

The metadata will be submitted to the archive of UoE, i.e. DataShare in a format relevant to their metadata requirements / standard.

Where will the data be stored and backed-up?

All electronic data will be stored in a password protected computer. Access to this computer will be only limited to the essential team members.

Data will be backed up on a shockproof external hard drive. This hard drive will be kept under lock and key at AAIP. Moreover, Backup of the data will be taken on the DataStore through DataSync. This will also have the advantage that the data can be reviewed and discussed with the team at UoE. We will not possess any type of sensitive information about the patient, not even the gender or age of the patient. Hence, once the project is completed, data can be shared on Edinburgh DataShare.

3. Integrity

How will you quality assure your data?

Overall, the quality will be assured by the PI who will oversee all aspects of the data collection, data entry and analysis.

Since this is a laboratory-based study, all steps of laboratory quality assurance will be followed as recommended by the World Health Organisation. The outcomes of the

laboratory procedures will formulate the data regarding the development of the diagnostic kits.

The laboratory quality assurance (including quality control) will be overseen by the laboratory manager and the co-investigators. The PI himself has remained a WHO Consultant on Quality Assurance in diagnostic laboratories and is very well versed with all related procedures.

The quality of the data transfer will be assured by a multi-step procedure in which each step will be double entered in Excel files, by two different operators at different times on different computers, and then cross checked periodically by the PI and the Data Manager.

The final quality assurance protocols will be followed under advice of RESPIRE, UoE.

4. Confidentiality

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

Approval of sponsorship has been obtained from ACCORD, furthermore, local Pakistani clearance has been obtained from the International Research Force.

All human participants in this study will be informed about the data being collected, the purpose of the study, and that the results might be shared with other researchers without compromising the privacy. Sensitive information like name and address, age and gender etc will be removed from the samples collected at the diagnostic lab, and the samples will be encoded before sending to the research team by giving each participant a unique anonymised identifier. This identifier will be a numerical value assigned sequentially from $1 \cdots n$, where n is the total number of participants in this study. Hence, the data will not contain any information that may directly or indirectly identify the participant.
The data generated in this project will be entirely co-owned by AAIP and UoE. However, anonymised data will be available to be used for research purposes.

5. Retention and Preservation

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

All anonymised and aggregated data will be available through Edinburgh DataShare indefinitely.

6. Sharing and Publication

Which data will be shared and how?

As previously discussed, both Edinburgh DataShare will be used for long term dissemination of the anonymised data. The data will always be first encrypted and then transferred to DataShare.

Are any restrictions on data sharing required?

During the term of project, the data will only be accessible to the data team and Dr. Osman (Project PI). For later research purposes, it will be shared according to the term of usage prescribed by Dr. Osman.