improving global
RESPIRE
respiratory health

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

Name:	Dr. Tabish Hazir
Project Title:	COVID-19 Detection from Chest X-Rays using Deep Learning
Institute:	Maternal, Neonatal and Child Health Research Network (MNCHRN)
Start Date:	15th Dec. 2020
End Date:	31st July 2021
DMP version number and date:	DMP_V01_20201130

Responsibilities & Resources (applicable across the sections below)

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

Name	Role in the Data Management
Dr Hana Mahmood	Data deposit and oversight of implementation of the plan
Syed Yahya Sheraz	Data storage
Nabit Bajwa	Data management
Dr Misbah Durrani	Data collection
Sajid Khan	Data collection

1. Data Capture

What data will be generated or reused in this research?

Chest X-rays will be obtained from both open-source public datasets as well as from the radiology department of Benazir Bhutto Hospital (BBH), Pakistan.

Open-source datasets being used are NIH CXR8 Dataset, Stanford CheXpert Dataset, and COVID Chest Xray Dataset.

From BBH around 500 X-rays of those individuals will be included in the study who:

- Will have either a PCR labelled COVID-19 diagnosis
- All age groups
- Able to provide consent for their X-rays to be used in the study. For children consent will be taken from the caregivers of those children

www.ed.ac.uk/usher/respire
@RESPIREGlobal
RESPIRE@ed.ac.uk

The chest X-rays will be obtained in the electronic form from the radiology department of BBH along with the details of the age, gender and presenting complaints of the patients.

Polymerase chain reaction (PCR) will be used as a gold standard diagnostic tool for labelling COVID-19 positive and negative patients. Once obtained, AI team of MNCHRN will use these X-rays for developing an artificial intelligence-based model for diagnosing COVID-19 which will be able to differentiate between X-ray of a COVID-19 to a non-COVID-19/normal individual.

How much data will be generated?

Up to 25-50GB of data will be generated

2. Data Management

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

Two databases will be created, one will contain the chest rays of all the patients and the other will contain the details (age, gender and presenting complaints) of the patients in an Excel workbook. X-rays and details will be given the ID assigned to the patients.

Where will the data be stored and backed-up?

Storage:
Data including chest X-rays, age, gender and presenting complaints of the patients will be stored in password protected and encrypted computers with access to authorized members of the research team only. These computers have been encrypted using Bitlocker and the folders are stored in encrypted containers using the 7-zip.

Backup:
3 backup copies of the original data will be created and stored on different locations. Original copy will be stored on the password protected and encrypted computer of the data manager placed at the corporate office of MNCHRN. One copy will be backed up on the main server located at the central field office of MNCHRN and two copies on cloud storage (My Cloud EX4100 secured by volume encryption with enterprise-class WD Gold Hard Drives- 2 drives of 10TB each) on fortnightly basis. These hard drives provide up to 2.5M hours MTBF, vibration protection technology, RoHS compliant, have SATA 6 Gb/s interface, 262 MB/s data transfer rate, 256 MB cache, 7200 RPM performance class.

File naming:
Files will be named as per the standard naming convention of our organization i.e., organization name_projectID_subjectID_type of data file

Organization:
Folders according to the type of data will be created and all files named according to the subject ID will be stored in it.

www.ed.ac.uk/usher/respire
@RESPIREGlobal
RESPIRE@ed.ac.uk

3. Integrity

How will you quality assure your data?

Following measures will be taken to assure the quality of the data:
- All versions of data collection tools will be saved for traceability.
- Data stored in three different locations to prevent the loss of the data. Original copy will be on the password protected and encrypted computer of the data manager placed at the corporate office of MNCHRN. This will be backed up on the main server located at the central field office of MNCHRN and also on cloud storage on fortnightly basis.

4. Confidentiality

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

Names and identifiable information will be removed to make them anonymous.

Additionally, consents will be taken from the participants indicating their voluntary participation and permission to share the data as per the participant information sheet.

5. Retention and Preservations

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

All data generated from this study will be kept for a minimum of 5 years after the study is finished.

How will the data be preserved?

The anonymised and cleaned data including X-rays and patient's presenting complaints will be preserved in stable file formats i.e., comma-separated value (.csv) file for spreadsheets. These will be uploaded in University of Edinburgh's online digital repository – DataShare.

The University of Edinburgh undertakes to maintain the digital outputs of this project for the long-term and will utilise University infrastructure (namely the Edinburgh DataShare repository) to ensure preservation and continued access. Edinburgh DataShare is an online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh, hosted by the Data Library in Information Services. It acts as a trusted repository, ensuring that research data will be preserved.

6. Sharing and Publication

www.ed.ac.uk/usher/respire
@RESPIREGlobal
RESPIRE@ed.ac.uk

Which data will be shared and how?

The anonymised data generated from this project will be shared and archived in DataShare which is an open access data repository where data will be held indefinitely. The repository is optimised to be indexed by Google scholar and Google dataset search to promote data discovery.

Are any restrictions on data sharing required?

No, there are no restrictions on sharing of the anonymised data shared and archived in DataShare.