



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

Name:	Dr. Nomana Zeeshan
Project Title:	A pedagogical research on use of an online learning platform by final year medical students on under five pneumonia
Institute:	Maternal, Neonatal and Child Health Research Network (MNCHRN)
Start Date:	1 st Apr. 2019
End Date:	30th Sep. 2020 (with 3 months no cost extension)
DMP version number and date:	DMP_V01_20201130
<u>Responsibilities & Resources (applicable across the sections below)</u>	
<i>Who will be involved in the data management of this research?</i>	
Name	Role in the Data Management
Nomana Zeeshan	Data collection, management and implementation of the plan and administration of the online portal
Hana Mahmood	Data deposit
Hira Kiani	Data collection
Saman Mujeeb	Data collection and maintenance of data
Sajid Khan	Data collection
Nyla Muzaffar	Data collection
Madeha Shah	Data collection
<p>Data have been generated and collected through the online learning portal developed as a part of this study for identification and management of pneumonia in patients under the age of five.</p>	

1. Data Capture

What data will be generated or reused in this research?

Registration data including the name, age, and contact details of the participant (final year medical students of five medical universities within Rawalpindi and Islamabad cities of Pakistan) were collected by the members of the research team, at the time of their enrolment in the study. Members of the research team collected this data on the hard copy of the registration form which was used to create their accounts on the portal. Each user was assigned an ID at this stage that included an alphabet to refer to the institution followed by a numeric code that was unique for each user. Users were sent login details to their email IDs.

After logging in to the portal for the first time, participants completed the following questionnaires related to pneumonia in patients under the age of five

1. Pre-test self-assessment of knowledge on pneumonia in patients under the age of five
2. Pre-test self-assessment of skills, confidence, and competence on diagnosing and managing cases of pneumonia in patients under the age of five.

Participants solved three case scenarios each month. The scoring of each case has been stored in the database. A total of 18 cases were presented to the participants over 6 months.

At the end of the intervention, participants completed the post-test questionnaires on self-assessment of knowledge, confidence, and competence in identifying and managing pneumonia in patients under the age of five. They also filled a feedback form regarding their experience, design, and content of the portal.

All of the data have been generated and collected electronically through the e-learning portal and stored in databases as excel spreadsheets with user's unique ID thus all the data in the database is de-identified. This data was then imported into SPSS for statistical analyses.

Only the administrator of the portal has access to the participant data and scores of each individual participant to be used for reporting. The access to this information is password protected and only accessible to the administrator.

How much data will be generated?

Up to 20 GB of data will be generated.

2. Data Management

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

Each user was assigned a unique ID which was used to track his/her score. The scoring of each case was saved in individual file so 18 files for test scores have been created. In addition to these, the answer key for each case scenario have been saved in excel file. Following details of all these excel files have been recorded in a text file:

- name of the file
- file location
- file size

- file format
- software used or required to create or run the file (e.g., Microsoft Word for transcripts and translated documents)
- date of creation
- file creator
- file version
- access rights set for the file

Where will the data be stored and backed-up?

Storage:

The score of each case scenario was downloaded from the database of the website and stored as a separate excel file in password protected computers with access to authorized members of the research team only. These computers have been encrypted using the Bitlocker and the files are stored in encrypted containers using 7-zip.

Files have been named as per the code assigned to the case, e.g., month1-case01, month2-case02 where the first part indicates the month followed by the number of cases of that particular month.

Backup:

3 copies of the original data have been created and stored on different locations in the form of excel files. Original copy has been stored in the database of the website hosting. Its copy was downloaded and saved in password protected computer of the data manager placed in the field office. It has been backed up on the main server located at the central field office of MCNHRN and on cloud storage (My Cloud EX4100, secured by volume encryption, with enterprise-class WD Gold Hard Drives- 2 drives of 10TB each) on a 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 have been named as per the standard naming convention of the organization i.e., organization name_projectID_subjectID_type of data file e.g., MCNHRN_RES004_I.1.1_transcript.docx

3. Integrity

How will you quality assure your data?

Following measures have been taken to assure the quality of the data

- Data stored on 3 locations to prevent the loss of the data as indicated in the Storage section of Data Management
- 2 factor authentications to verify the user on first login.

4. Confidentiality

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

By ensuring the anonymity of the data by removing identifying details or giving pseudonyms. The data have been de-identified as it has been saved with unique user IDs and not the actual names. The results will be quantitative, and analyses is being done through SPSS. The main aim of analyses will be to see the effectiveness of the portal and will not involve any form of ranking the students or their institutes.

Additionally, consents have been taken from the participants at the time of their enrolment in the study.

The IT security aspect of the portal is as follows:

- a) Data will be generated and stored in a database which resides on *nix platform.
- b) Login user passwords will be stored in SHA-2 hash.
- c) Anti-DDoS mitigation.
- d) Firewall that opens https and custom ssh port to public.
- e) Password strength enforcement with upper, lower, number and symbols.
- f) SHA-2 and 2048-bit encryption.
- g) Secure socket layer
- h) Data transmission over SSL https protocol with 2k encryption.

5. Retention and Preservation

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

All data generated from this study will be kept for a minimum of 3 years after the study is finished. Answer keys of all scenarios, and the pre and post-test questionnaires will be preserved as well.

How will the data be preserved?

The de-identified and cleaned data including the pre and post-test questionnaires, pre- and post-test scores, scores in the 18 case scenarios, answer keys of all the case scenario, the analyses of scores and meta data will be preserved in stable file formats i.e., note pad for text files and 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

Which data will be shared and how?

The anonymised data generated from this project will be shared and archived through DataShare which is an open access data repository.

Are any restrictions on data sharing required?



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

