



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

Name:	Dr. Anand Kawade
Modified by:	Tathagata Bhattacharjee
Project Title:	ASHA's workload perception and its' determinants
Institute:	KEM Hospital Research Centre, Pune, India
Start Date:	01 October 2018
End Date:	30 June 2019
DMP version number and date:	V1, dated 02 October 2020

### **Responsibilities & Resource (applicable across the section below)**

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

ASHA stands for Accredited Social Health Activist. For efficient data management of ASHA study, different categories of resources were involved, inclusive of human and other types.

The roles of human resources involved were:

1. Field Research Assistants (FRA) for on the field work
2. Field Coordinators (FC) for aligning field activities and back-office data entry
3. Data Entry Operators (DEO) for transferring the paper form data into electronic format
4. Senior researchers and Co-investigator for overall coordination and management of the study
5. Data Manager (DM) for electronic data management
6. Independent Qualitative researcher for translation and analysis of qualitative data
7. The Principal Investigator (PI) for overall responsibility of data generation

The individuals assigned the specific roles described above were:

1. Field Research Assistants (FRA) (Amol Bhosale, Meera Tambe, Shalan Mhetre, Sharda Chaudhary and Padma Neherka) for quantitative data collection on field and supported the qualitative data collection
2. Field Coordinator (Pradnyand Jondhale) responsible for supervision of data collection, quality check and coordinating data collection activities on field and data entry activities at office level.
3. Data Entry Operators (DOE) responsible for data entry of collected data.



4. Senior researchers (Uddhavi Chavan, Pallavi Lele) were responsible for In-depth interviews and key informant interviews and were supported by field team
5. Co-investigator (Pallavi Lele) was responsible for overall coordination of the study
6. Independent Qualitative researcher (Prashant Kulkarni) was responsible for translation of qualitative data.
7. Independent Qualitative researcher (Manisha Gore) was responsible for analysis of qualitative data supported by senior researchers

Other resources utilised for data management processes during this study include the following:

1. Computers (desktop and laptops) used for data entry work.
2. External hard drives for backup during the data entry work.
3. Backup servers for storage of data at Vadu site, KEMHRC and cloud services.

At the end of this project data will be submitted to Edinburgh DataShare (<https://datashare.is.ed.ac.uk/>) for sharing data in public domain and for long term preservation on DataVault (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/datavault>)

## 1. Data Capture

### *What data will be generated or reused in this research?*

The data generated in ASHA study is summarised as follows:

#### **Questionnaire data- (quantitative data) in CSV format:**

The quantitative data were collected using paper-based self-administered questionnaires. All ASHAs responded when approached during their routine monthly meetings and vaccination camps. The closed questions were on socio-demographic profile, time spent on ASHA work and travel, perceptions of workload and its impact on them and their family, opinions about remuneration, job satisfaction and family support.

#### **Qualitative methods:**

Qualitative data were collected using in-depth interviews. We purposively sampled eight ASHAs, four from each of the Rural and Tribal Primary Health Centres (PHCs), based on diversity of experience (> 5 years), educational background and other paid work. We also interviewed the two Medical Officers (MO), four Auxiliary Nurse Midwives (ANM) and four Block Facilitators (BFF) from the same PHCs, each of whom had at least a year's experience of supervising ASHAs.

We had developed a conceptual framework to inform the interview guide based on open-ended informal discussions with ASHAs and other colleagues. This helped us to understand the workload in terms of time investment, travel, energy and effect of the work on ASHA's family

and self, and highlighted age, training, education, experience, work setting, incentives and other occupation as influencing factors. The volunteer status of ASHA was important in interpreting relative prioritisation of their work.

Topic guides were prepared in English and translated into the local language (Marathi) by the Field Research Assistants during their training, reviewed by researchers, and piloted before finalisation. During data collection, the conversation focused on enquiry around typical daily activities. This led to probing on ASHA activities, household activities and other occupations, their perceptions of workload related to the different tasks, the challenges and motivations to continue the work. The qualitative interviews were conducted by graduate-level trained qualitative researchers, supported by a note-taker who had experience of working with the health and demographic surveillance system.

The above-mentioned data were collected in hard copy forms and would be destroyed after a minimum five years from the protocol defined end of study point. Fully anonymized (without any identifiers) soft copy data are being stored by KEMHRC as per its data sharing and archiving policy which is in line with the guidelines set by the national (Indian) data access and sharing policy.

***How much data will be generated?***

Total 90 records were generated for quantitative and qualitative data type mentioned in above section. The size of data is less than 1 Gigabytes.

**2. Data Management**

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

Each type of data as stated in section above were collected using well-structured questionnaires. The questionnaires had instructions with pre-question, the literal question and post-question information sections. These questionnaires (blank templates) both in soft (pdf) and hard copies are stored at site along with the soft copy datasets and hard copy filled forms respectively for any future references.

All hard copy data with respect to this study and its documentation, inclusive of codebook wherever applicable are kept for a minimum of five years from the protocol defined end of study point.

Data quality control was done using Stata v15 tool and appropriate labels were assigned wherever necessary for manual quality monitoring and checks. However, in the final data formats, which are csv and text, the labels were not included.

All interview data were collected with the help of FRA and qualitative researcher. The interviews were guided by conceptual framework. The interview data were transcribed into

Marathi text by FRA which further translated into English by independent qualitative researcher. Senior researcher checked the quality of data before entering it in MAXQDA for analysis

It is planned to do proper Data Documentation Initiative conforming to international standard for describing surveys, questionnaires, statistical data files, and social sciences study-level information. This added detailed and quality metadata for the datasets generated in this study. Metadata documentation is planned to be done for study datasets. The plan for this data documentation included complete study documentation along with all the processes and standards incorporated and adhered to along with the other data metrics as was identified during the process. A detailed variable level metadata will be created for easy end-user understanding at any point of time.

### ***Where will the data be stored and backed-up?***

Data was stored and backed up as per the below listed principles:

1. All data generated in relevant retrospection, joint studies and partnered projects under RESPIRE may have a cleaned and anonymized subset copy on the UoE data repository, named as DataShare. Access to such data on DataShare will be public.
2. A copy of all data that is uploaded on DataShare is retained by KEMHRC on an “as-it-is” basis along with a master mapping record for identifiers. This is needed for regulatory purposes. The copy retained at KEMHRC will not be uploaded on any other public access data repository unless agreed by both the UoE and KEMHRC.
3. All data generated in relevant retrospection, joint studies and partnered projects under RESPIRE may be put in Edinburgh DataVault, for long term preservation, however the copies on DataVault must be anonymized with master mapping data for the identifiers in custody of KEMHRC. Dataset on DataVault must have controlled access with a definite lifetime assigned as per institution’s policy.
4. For all datasets pushed on to DataVault, a copy must be retained by KEMHRC with assigned lifetime as per institution’s policy (5-8 years for KEMHRC) along with the master mapping data for identifiers. The location of storage and related services will solely be the responsibility of KEMHRC.
5. All in-process data, i.e., active research data, that may need sharing with group members remotely may be put on UoE’s DataStore (<https://www.ed.ac.uk/information-services/research-support/research-data-service/during/data-storage>). These types of data sharing will be guided by the MoU and data sharing agreements of the collaborating institutions.
6. KEMHRC’s document server may also be used for all in-process, i.e., active research data that needs sharing whilst working collaboratively within office premise local network or VPN.

7. All data on either DataShare, DataStore or on DataVault, the ownership lies with KEMHRC with grant of custody given to UoE under terms and conditions of MoU.

Based on the above principles, data generated for the ASHA study was stored as described here:

1. The data storages of KEMHRC includes the following and all data stored was catalogued using standard methods and are considered as “enclaved”, meaning that no direct access would be given. Probable users can search from the catalogue and raise a request for copy of the data.
2. KEMHRC data storage server is located in Pune office. This storage server is a well configured secured storage for all project data and catalogued and accessible over local network only. These are not publicly available resources and are accessible from within the network in office premises.
3. The KEMHRC data storage server is also configured to serve as a document server and all in-process, i.e. active research data, that needs sharing with group members can be used for access from with the local network or over VPN.
4. KEMHRC data storage server located at Vadu office. This is a temporary storage server for storing in-process data and does not store the final archival versions and accessible over local network only. These are not publicly available resources and are accessible from within the network in office premises.
5. A complete copy of raw data permanently archived in the above-mentioned KEMHRC data storages and catalogued for a minimum period of eight years in order to comply with the KEMHRC data policy, IT laws of India and funder/sponsor requirements.
6. In-process data, i.e. active research data, if needed, may be put on UoE’s DataStore (<https://www.ed.ac.uk/information-services/research-support/research-data-service/during/data-storage>) in cases of distributed teams to share files anywhere and with anyone with study groups.

### **3. Integrity**

#### ***How will you quality assure your data?***

Quality checks of the questionnaire data were done at three levels.

1. FRAs involved in data collection checked the collected data for completeness and logical checks. Thereafter, once data was received at office, field supervisors perform quality checks by using pre-defined criteria. Data showing issues were sent back to the field by the supervisor through reassigning those to the respective FRA. Field coordinator performed random quality checks for a few questionnaires.
2. The cleaned data were entered in excel sheet by FRA. It was further checked logically by senior researchers (Uddhavi Chavhan & Pallavi Lele) and further entered into STATA for analysis.

3. All the recorded interview data were transcribed by FRA into text in local language. Further it was translated from local language to English by qualitative researcher (Prashant Kulkarni). The translated data were checked for quality by independent researcher (Manisha & Uddhavi). The final cleaned data were entered into MAXQDA for analysis.

#### **4. Confidentiality**

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

All Investigators and study site staff involved with this study conformed with the requirements of the General Data Protection Regulation (GDPR) 2018 with regard to the collection, storage, processing and disclosure of personal information and uphold the Act's core principles. Access to collated participant data is restricted to individuals from the research team, treating physicians of the participants, representatives of the sponsor(s) and representatives of regulatory authorities.

Computers used to collate the data and have limited access measures via usernames and passwords.

All identifying information that is collected about the participant (such as name, age, sex, address, contact information) during the course of the research is kept confidential and secured. Published results will not contain any personal data that could allow identification of individual participants.

The data of each study participant will be identified with the help of a unique identifier and it will be completely anonymized and scrambled before sharing. The details of the unique identifier will be held with the research team. There will be no such information in the shared data which will disclose the identity of the study participant. Standard and recommended security measures and confidentiality with data sharing agreements were in place with access control at every stage and audit trails maintained for all access and changes in data.

#### **5. Retention and Preservation**

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

Data is retained and preserved as per the principles stated in section above in Data Management.

All hard copy data (filled forms), which includes identifiable information and related documentation is preserved at KEMHRC Vadu for up to a period of five years from the protocol defined end of study point. After the elapse of five years, hard copy data will be destroyed as per KEMHRC guidelines and/or specific contract clause with the sponsor(s), if any or under prevailing law of the land (India).

Soft copy of the raw data is uploaded on secured KEMHRC data storages with limited access to KEMHRC data administrators only. Data on KEMHRC data storages are catalogued. Data is “enclaved” in the storages, meaning it is findable through the catalogues but no direct access is given. Data is categorized and some categories of data, for example the identifiers, which are for internal reference only will not be made accessible to non-KEMHRC entities. The categories of data meant for public access either open or controlled will not be on these storages.

Any access needed is to be directed through the data administrator after due approvals. As per KEMHRC policy, this soft copy of data will be retained on the storage server(s) for a minimum period of eight years with no upper limit defined.

An anonymised copy of the study data will be backed up on the UoE’s DataVault (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/datavault>) for long term preservation. The preservation details are articulated under the next heading.

### ***How will the data be preserved?***

Based on the principles listed in section 2, data generated for the said study is/will be preserved at the end of study as described here:

1. Soft copies of all data collected in this study is anonymised with identifier mapping master.
2. Soft copies of all data is preserved by KEMHRC along with the mapping master was retained as per data policy of KEMHRC (The KEMHRC data policy is not made available as public accessible resource as on date; however, it is sharable with collaborators on approvals from the trust members).
3. Data is preserved on University of Edinburgh’s DataVault (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/datavault>) for a longer period as defined by the University’s data policy.
4. KEMHRC has preserved data on its data backup servers located in KEMHRC Pune office and also on commercially purchased data archival cloud space (<https://aws.amazon.com/glacier/>).
5. All soft copies of data including identifiable information and related documentation is preserved on KEMHRC storages and anonymised copies on UoE’s DataVault (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after/datavault>).

A complete copy of the anonymised data validating the results is preserved for long term in the above-mentioned data storages and catalogued for a minimum period of eight years in order to comply with the KEMHRC data policy, IT laws of India and funder/sponsor requirements.

## **6. Sharing and Publication**

### ***Which data will be shared and how?***

Data sharing principles encourage ethical commitments of data generated from the public and must benefit the public by sharing for open access research opportunities. KEMHRC holds this principle to its core and provides data from its studies and projects for sharing after due processes of cleaning, anonymisation and masking confidential information wherever applicable.

For this study, KEMHRC submitted anonymised data for public access on University of Edinburgh's DataShare (<https://datashare.is.ed.ac.uk/>). Data on DataShare must follow the principles of findability, accessibility, interoperability, and reusability (FAIR) and the submitted dataset must have a Digital Object Identifier (DOI) assigned.

As per KEMHRC policy, data on KEMHRC server will be stored for a minimum eight years with no upper limit defined. A similar copy of the data would be retained by KEMHRC for adherence to local IT laws. Any derived or calculated or other form of data can also be shared on DataShare.

### ***Are any restrictions on data sharing required?***

There are a few restrictions and procedures for compliance to KEMHRC data policy and local IT laws:

- (1) Identities of study participants cannot be shared or stored on servers outside the boundaries of India
- (2) Only anonymised data can be shared on public domain. The degree to which anonymisation is done must be clearly understood and documented.
- (3) A copy of all data stored on servers outside India must have a copy within Indian territory and must be made available to any law enforcing or regulatory agency on demand
- (4) The law enforcement and regulatory authorities will have full access to the data as per the rules and regulations.

Not all of the above is law yet but compliance is solicited. It is expected that any researcher using this dataset for any type of publication or conference paper must cite this dataset by referencing the DOI.

====end of the document====



[www.ed.ac.uk/usher/respire](http://www.ed.ac.uk/usher/respire)  
[@RESPIREGlobal](#)  
[RESPIRE@ed.ac.uk](mailto:RESPIRE@ed.ac.uk)



THE UNIVERSITY  
of EDINBURGH

FUNDED BY  
**NIHR** | National Institute  
for Health Research

