PUBLIC HEALTH FOUNDATION of INDIA

DATA SHARING POLICY

REVISION HISTORY		Approved On
Latest Version	V 4.0	14/11/2020
Superseded By	V 3.0	02/09/2017

Approved By:

Dr. D Prabhakaran

Signature:

Dated 14/01/2021

Table of Contents

1. INTRODUCTION 3

1.1 DEFINITION 3

1.2 OUR GUIDING PRINCIPLE 4

2. RATIONALE FOR DATA SHARING POLICY 4

3. GOALS OF DATA SHARING POLICY 5

4. SCOPE AND BENEFITS OF DATA SHARING POLICY 6

5. TERMS OF AGREEMENT FOR DATA SHARING AND ACCESS 6

5.1 FOR THE PI/ DATA SOURCE AGENCY: 6

5.2 CONDITIONS OF NON-EXCLUSIVE USE OF DATA SETS FOR THE DATA USER: 7

6. DATA ACCESS AND SHARING- REVIEW OF REQUESTS 7

6.1 TIMELINES OF DATA SHARING 8

7. Central Research Data Repository 8

7.1 BENEFITS OF REPOSITORY: 9

7.2 AVAILABILITY OF DATA FOR INTERNAL AND EXTERNAL RESEARCHERS: 9

8. UPDATING/ REVIEWING THE POLICY 9

9. MANDATE FOR NEW PROJECTS 9

10. RATIFICATION OF THE POLICY 10

11. HANDLING OF BREACH OF THE POLICY AGREEMENT 10

12. LEGAL ISSUES PERTAINING TO DATA SHARING 10

13. ETHICAL ISSUES PERTAINING TO DATA SHARING 10

14. FREQUENTLY ASKED QUESTIONS ON DATA SHARING 11

15. REFERENCES 17

16. ANNEXURES 18

Annexure I: Data Access Flow Diagram 18

Annexure II: Data Access Variables/Questions 20

PART A: Basic Details required for Data Access 20

PART B: Confidentiality agreement Form for Data Access and Usage– PHFI 21

Annexure III: Summary Of Funder's Requirements For Data Sharing 22

Annexure IV: Publisher's Requirement 28

Annexure V: Format Data use Agreement – PHFI 29

Annexure VI: Repository Management Information Layer Application Features: 33

Dated 27/11/2020

Summary

Institutional Data Sharing Policy

This document encompasses the ways in which data derived from research, operational and implementation programmes (data) once accrued at the Public Health Foundation of India (PHFI) and its constituent Indian Institutes of Public Health (IIPHs), may be stored, archived and shared.

This policy is specific to PHFI meant to safeguard the investigators and team members involved in the collection of data and the organization. At PHFI, the Research Management Committee (RMC), Internal Technical Review Panel, Central Research Data Repository Cell, Principal Investigators and the relevant project team staff are entrusted with managing and protecting the data arising from various projects undertaken. This policy applies to all projects that are undertaken at and by PHFI.

This will complement those for intellectual property rights, institutional ethics approval and financial protection at PHFI/IIPHs and will not supersede them. Guidelines or policies that are drafted to manage authorship criteria will be independent of this document.

I. **Definition**: Data pertains to records, information, numerical raw results, and outputs from instruments, survey questionnaires, notebooks, electronic or paper documentation of experiments, data generated from trials, data analysed from raw results, data collated and systematically represented during the course of an experiment.

II. This policy is applicable to all PHFI/IIPH staff, researcher, faculty, students and consultants who are involved in projects/programmes undertaken by and at PHFI/IIPHs and aligned to as per the requirement of funding agencies.

III. The policy will be applicable for all the projects by the date of approval of this policy. For new studies as per the requirements specified by the donor, all the documentation will be carried out in repository.

IV. Ownership of the data generated by projects conducted at or within the auspices of the organization, rests with PHFI and Principal Investigator (PI).

V. In general, irrespective of the data management systems used by the research teams of PI, the PI would select an appropriate method for storing and archiving data generated from respective project during the initial phase of the study. The PI will also be responsible for educating the team members pursuing the respective project to follow the data storage and archiving methods. These methods should ensure that data is accurate, authentic and adheres to compliance if the organization has to respond to any queries regarding the data at any date, during and after the project is completed.

VI. The PI will have exclusive access to the data generated by the respective project and to others that are granted permission by the PI. Team members will have a right to review that portion of the data that was generated/created by them. PHFI will have access to the data as necessary for compliance and other purposes. In case the sponsoring agency/collaborators wishes to access data generated by the respective project, this will be based on the terms and conditions and policies (MOUs/agreements with the donor and collaborators, PHFI corporate policy and conflicts of interest policy) defined and followed by the PI and the collaborator/sponsoring agency.

VII. For technology transfer, intellectual property rights, royalties based on the data generated; the PI, primary contributor from the team, legal, research admin and finance will create a separate agreement that will be reviewed and approved by the Research Management Committee following the norms and regulations of PHFI as applicable. Wherever applicable by law/agreement, access to the data would be provided to the donor or collaborators of the study or to the journal publishing the data by the PI and their team.

VIII. Data sharing: Requests to share the data from collaborators or external agencies will require prior approval on a separate data sharing and management form. These forms must be pre-reviewed by the Internal Technical Review panel of PHFI, comprising of some Research Management Committee Members, Senior Management team members and core CRDR team. Those researchers, getting funds from other institutions may have financial undertakings for data sharing expenses that will be reviewed by finance and legal department of PHFI. These forms would conform to the present data management and sharing guidelines, if any with respect to the individual project. Final approvals will be provided by the above mentioned Internal Technical Review Panel of PHFI and in special cases, by the Research Management Committee, which will then be coordinated by Head, Central Research Data Repository. PHFI/IIPH and its PIs and applicable team members should be appropriately cited in any publication derived from the data that is shared by the organization.

IX. Data will be maintained by PHFI for at least seven years after completion of the project or as specified by the funding agency. Appropriate scanned copy of paper and electronic storage mechanisms would be made so that data can be accessed as required. Data will not be destroyed or removed without prior approval of the Research Management Committee.

X. Transfer or removal of data from PHFI to another organization after the seven years period in point IX; must be approved by RMC. PHFI would retain copies in all such cases.

XI. After completion of the project, the PI can store and archive the data in the PHFI central database repository.

XII. Any conflicts emerging from the data sharing/access process at any stage, would be dealt with by the Internal Technical Review Panel constituted by the Vice-President, Research and Policy of PHFI and further by the RMC wherever required.

XIII. In the event of PI changing/leaving PHFI, a copy of the research and its data should be transferred to PHFI. The PHFI - IT will provide the NOC to the respective PI, post transfer of the data to the CRDR repository.

XIV. Detailed Process Flow chart is given for ready reference to all the PI's and the Researchers. (Annexure 1)

1. INTRODUCTION

Studies of health investigators focus on health care. Data sharing help to put research into practice i.e. new/alternative treatments and research knowledge benefit the populations for whom the studies are conducted. It also helps in health decision making. Moreover, it increases the impact and visibility of research. It promotes innovation and potential new data uses. Basic research, clinical studies, surveys, and other types of research produced data may be shared. A data repository achieves the goals of data sharing.

PHFI is committed to ensuring that the outputs of the research it conducts, including research data, are managed and used in ways that maximizes the public benefit. It will adopt policies that aim to increase the use and reanalysis of the data. It also expects researchers to state whether and how they will make protocols, analysis tools and data available to others on request. Making research data widely available to the research community in a timely and responsible manner ensures that these data can be verified, built upon and used to advance knowledge and its application to generate improvements in health. It will speed up the research that may eventually prevent or treat the disease. Data sharing is particularly important for the unique data that cannot be readily replicated. Data sharing allows the scientists to expedite the translation of research results into knowledge, products and procedures to improve human health. The PHFI is developed a policy on data sharing that expects and supports the timely release and sharing of final research data from PHFI supported studies and studies done through other grants by the PI's of PHFI for use by other researchers. The Central Research Data Repository will be based out in Plot No. 47, Sector 44, Institutional Area, Gurgaon - 122002, Haryana.

1.1 DEFINITION

Terms	*Meanings*
Funder	The name of the funding agency to carry out the research work.
Data sharing rights	Funder statement on license terms that should be applied to research data made available
Data Standards	General or specific data standards or practices that should be adopted by the researcher when creating, managing, or sharing data. For e.g. please refer to the data quality mentioned at https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002056
Data Plan Requirements	Requirements to provide information on data management and/or sharing activities as part of the proposal to be submitted.
IPR	The term "Intellectual Property Rights" refers to the legal rights granted with the aim to protect the creations of the intellect. Intellectual Property (IP) refers to the protection of creations of the mind, which have both a moral and a commercial value.
Retention Requirements	An indicator of the minimum time period that data assets should be retained and the institutional agent that should be responsible
Meta data	Meta is a prefix that in most information technology usages means "an underlying definition or description." Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier.
Documentation Requirement	Documentation suggested by funders that should be provided to help researchers to access and use research data produced.

Terms	*Meanings*
Central Database Repository	A Central database repository is a logical and physical grouping of data from related but separate databases with built in-depth access control rules.
PHFI/IIPH	Public Health Foundation of India/Indian Institute of Public Health
Relational Database Management System (RDBMS)	RDBMS data is structured in database tables, fields and records. Each RDBMS table consists of database table rows. Each database table row consists of one or more database table fields. It stores the data into collection of tables, which might be related by common fields (database table columns). RDBMS also provide relational operators to manipulate the data stored into the database tables.

1.2 OUR GUIDING PRINCIPLE

Being evidence centred and ethical – to work with new and existing data from a variety of sources to generate knowledge that is translated to research based intelligence.

- ✓ To support development of healthy sustainable future by building a rich body of knowledge, insight and commitment.
- ✓ To encourage fresh thinking and new approaches.

Data sharing policy is to facilitate the practice of making data generated through research available to other researchers, data users, funding agencies and publication readership.

PHFI/IIPH has developed this data sharing policy to ensure that our research builds on previous efforts and discoveries, that data access is ethical and in compliance with funder, publisher and Indian government agency requirements [1,2,3,4,5] and to maximize the potential and exposure of our findings to the larger scientific community, to policy makers, to health professionals, to community organizations and the public, at large. These are various channels for dissemination of data and mutual relationship allow learning to be acted upon. This is close and collaborative partner relationship that allows to two-way learning. This policy sets out specific procedures, timelines and access levels of data sets according to data sharing requirements of PHFI's all funding agencies.

2. RATIONALE FOR DATA SHARING POLICY

All data collected in the course of PHFI projects is the property of the institution, but under the control of the respective PI as long as they are at PHFI. Researchers have an obligation to make efficient use of this data to facilitate and stimulate exchange of ideas, new line of thinking, support processes of development. In public health research, however, while research collaborations are growing more common, the sharing of data is not yet the norm, even within the scientific community. Much of the data collection that could improve public health research is expensive and time-consuming. As a non-profit research institution, being funded by public and charitable organizations, we believe that making research data sets available to the scientific research community beyond the original research team inside as well as outside the institution in a timely and responsible manner, subject to appropriate safeguards, will generate these key benefits:

- ✓ Faster progress in improving health and Minimizes fund expenses, as duplication of collection of data avoided ensuring reusability of data.
- ✓ Ensures compliance of funder requirements and/or publisher regarding availability and access of final research data
- ✓ To reinforce open scientific inquiry and encourage diversity of analysis and opinion and validation of research methods and to test new or alternative hypotheses and methods of analysis.
- ✓ Sharing maximizes transparency and accountability and saves time of annotating data.
- ✓ Data comparisons among similar data sets encourage subsequent and cross study analysis.
- ✓ To promote new research, and help other researchers to build upon the existing ones not envisioned by the initial investigators and Facilitate two way learning that enhances the relevance of research.
- ✓ Sometimes data along with programs help in analysis for students and researchers specially.
- ✓ To permit the creation of new datasets when data from multiple sources are combined
- ✓ Sharing of data enhances reputation of Initial Investigator.
- ✓ Help to identify potential solutions and ways for improvement.

3. GOALS OF DATA SHARING POLICY

- ✓ All data should be considered for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data
- ✓ To provide a centralized platform and set of rules with greater access to data generated by PHFI.
- ✓ To maximize the number and type of analyses performed on valuable resources & to minimize the duplication of effort
- ✓ To create a standardized methodological and informational infrastructure on data collection procedures, protocol, cleaning, management and preliminary analyses and access to the database.
- ✓ To standardize, track and evaluate the flow of work generated from PHFI projects to publications.
- ✓ To track and evaluate the fund utilization
- ✓ Cash flow information like overrun etc.
- ✓ Centralized document management system with searching, capturing, storing and viewing version controlling research documents with the backup facility.
- ✓ Audit facility to keep the track of WHO, WHAT, WHEN, WHY has made the changes to the raw data with user login time, Activity and logout time.
- ✓ Well documented data sets are available for secondary analysis. Data collected for health research are made available to the scientific community for analysis which adds value to existing knowledge and which leads to improvements in health.
- ✓ Published work and data are linked and archived to the extent possible, Datasets underpinning research papers in peer-reviewed journals are archived and made available to other researchers in a clear and transparent manner.
- ✓ Long term sustainably resourced data sharing.

4. SCOPE AND BENEFITS OF DATA SHARING POLICY

This document provides an overview of suggested access and data sharing for PHFI projects and investigators, however different approaches to data sharing may be required according to the PI, a multi-Centre Project/Consortium, the funding agency (See Annexure I), a specific government policy and/or publisher specifications. In all instances, PHFI policy will attempt to adhere to the basic principle of data sharing as follows:-

- ✓ Data shall be made available for co creation, as quickly as possible.
- ✓ Information on frequency of data updates shall be provided in the metadata.
- ✓ Data shall be shared in machine readable format (xls, csv).
- ✓ The final dataset might include both raw data and derived variables, which would be described in the documentation associated with the dataset.

It applies to research data and related information collected and generated by PHFI in accordance with the data sharing plan submitted to funding agency at grant application stage.

This policy will strengthen PHFI capacity to effectively manage, curate, analyse research data and publish and disseminate research outputs as well as to foster strong and balanced collaborations and partnerships with the scientific/research community, within and outside the institution.

5. TERMS OF AGREEMENT FOR DATA SHARING AND ACCESS

It is possible to make data widely available to the research community while still safeguarding integrity, through the use of standardized data use agreements and licenses. These define who may use data and how, and may require secondary analysts to contribute both derived data and a record of their analytic methods back to the database, so that primary and other users can both verify and benefit from their work. It would be helpful for members of multiple disciplines and their professional societies to discuss data sharing, determine what standards and best practices should be proposed, and create a social environment that supports data sharing.

The researcher will ensure to keep a copy of his/her work on the data taken from the repository, back in the repository with the new version and synopsis for other researchers as a reference document. The Data license will be given as per the approval by the PI to the respective researcher of the PHFI/external researcher. The cost of the License/IT support will be borne by the project mainly. All the video files10 will be saved in MP4 and Audio files in MP3 along with Images/Pictures in JPG. All raw data files will be in .CSV, XLS, MS WORD and in PDF format. All media content will be available in the “DOWNLOAD as a file only” format to save the processing time of the server.

5.1 FOR THE PI/ DATA SOURCE AGENCY:

- ✓ Data released for sharing by the research group should be validated and verified in line with accepted best practices in the research field and be of high quality
- ✓ Shared datasets by the centre must be anonymized and should not contain data elements that will allow the direct identification of the entity (individual name and ID)
- ✓ The data sharing and access procedures should be transparent, and the data sharing review committee should have appropriate expertise.
- ✓ The funding agency and/or publisher may have specific requirements that override those declared by PHFI policy.

- ✓ The PI/ the source agency will not be responsible for any claims that the application of information generated from the data lead to wrong conclusions/decisions by the data user or other third parties.
- ✓ The PI should provide the data availability statement to the Research Management committee.

5.2 CONDITIONS OF NON-EXCLUSIVE USE OF DATA SETS FOR THE DATA USER:

- ✓ The data user shall submit a request for access using the specified format (Annexure II) and adhere to the PHFI guidelines, protocol and timelines on data sharing/access (diagram).
- ✓ The data user's application must include appropriate contact details, and a detailed description of the proposed methods, objectives and analyses, as well as any requests for biological samples.
- ✓ The access to biological samples should be restricted within PHFI and if any external agencies requested for this data for doing research work outside the organization/ outside India then the policies relating to transfer of biological samples within/outside the country should be adhered to. (An internal policy will be drafted for this by the research team dealing with the biological samples).
- ✓ The data user should protect the privacy of subjects and the confidentiality of the data as well as observe ethical and legal obligations pertaining to the data. A signed confidentiality agreement form should be completed by him/her to access the data set. In addition, the data user should secure the data using appropriate safeguarding techniques (see Annexure IV).
- ✓ All the data users who are accessing / using the data shall acknowledge the source agency and recognize contribution of data creator in all forms of publication. They should cite the source, in line with the citation requirement provided with the dataset.
- ✓ The Principal Investigator from the source agency should be given co-authorship in any research publication arising from the particular data set.
- ✓ The data user shall not transfer the data to other users in any form, must adhere to the ethical permissions/approvals, and must complete the work as per the formal request and approval of the data sharing review committee. Copying, modifying or adaptation of data sets to suit ones need. For this all the data security measures are in-built in the CRDR application.
- ✓ Must not use data that is inappropriate or misleading to general public i.e. using the data to promote any illegal activities or presenting datasets in a misleading manner.

6. DATA ACCESS AND SHARING- REVIEW OF REQUESTS

The Internal Technical Review Panel for CRDR will review all the proposals/ requests for data access. This Review panel would include personnel with appropriate expertise and the review would be based on the following:

- ✓ Scientific rationale and relevance of the proposed research
- ✓ Statistical analysis plan (design, methods and analysis)
- ✓ Personal/Professional details and research experience of the data user(s)
- ✓ Overview of funding sources for the proposed research
- ✓ Declaration of any real or potential conflicts of interest

The Internal Technical Review Panel should keep a record of data access requests which includes: data set(s) requested, date of application, applicant information including name, contact e-mail/address,

and institutional affiliation, relevant publications, whether the applicant's proposal/request was approved or not, date of approval/denial of access and reason(s) for approval/denial of access and level of access approved.

6.1 TIMELINES OF DATA SHARING

The suggested timeline starts after data lock which is after data collection and cleaning is complete and ready for analysis. The data would be available for access in three phases according to the interested data user/party:

For ongoing studies:

- ✓ **Phase I**: Within 2 years of data locking, access is provided to researchers/staff of the PI and his/her team (and collaborators) with data available via a shared folder /repository.
- ✓ **Phase II**: Within 3 years of data locking, access is provided to other researchers/staff within PHFI (with PI approval) with data available via the on-line repository of tools and/or shared on CD-ROMs.
- ✓ **Phase III**: After 3 years of data locking, access is provided to researchers/staff outside of PHFI with data available via the on-line repository of tools and data.

For new studies who opt for the CRDR platform:

- ✓ The data could be shared with in 2 months' time from the collection of data begins in the repository. This will help to manage the data more appropriately and can be shared with in the team of PI efficiently. (Refer – FAQ Q3 and Benefits of Repository)

For completed studies:

- ✓ For the closed studies, the cleaned data shall be shared with the repository team. For the ongoing & new studies the cleaning process can be executed using the CRDR platform itself.

These timelines may vary according to PI and collaborators, funding agency and/or publisher requirements.

7. Central Research Data Repository

A Central Research Database Repository is a logical and physical grouping of data from related but separate databases. This is usually done when there is a 'higher purpose' for the data, but the data items needed to do this reside on different databases. In these cases, a repository is necessary to bring together the discrete data items and operate on them as one.

- ✓ Research data intended for wider dissemination and thus submitted to a data repository that will curate, preserve and share research data after taking due permission from the PI who coordinates with the repository team to share the data with an external investigator.
- ✓ It is a managed environment capable of storing and sharing physical / digital data through host and query language.
- ✓ It provides stable, reliable and cost effective means for distributing data.
- ✓ Provide protections for the data set and technical assistance for the requesters.
- ✓ Allows for more than one version of the dataset and provide different level of access depending on the version.

7.1 BENEFITS OF REPOSITORY:

- ✓ To bring all research data collected from or available at different sources at a Central location for performing Trend Analysis and Predictive Analysis etc.
- ✓ Improved operational efficiency and productivity, security with increased consistency.
- ✓ Lower total cost of ownership
- ✓ To permit views by using various dashboards - on the data gives good trend analysis to understand the data in depth with various criteria mentioned in research protocol.
- ✓ Reduced data entry, reduced retrieval and storage cost.
- ✓ Reduced data redundancy by normalization.
- ✓ To ensure compliance with the funder's requirements (Annexure III): within the stipulated timeline in recognized data repositories/ institutional repositories maintaining high quality and standards.

7.2 AVAILABILITY OF DATA FOR INTERNAL AND EXTERNAL RESEARCHERS:

- **Internal (PHFI/IIPH)**
  - ✓ Repository of research Data, tools and datasets along with PDF files including informed consent forms, IEC approvals, protocols, protocol amendments, all project reports, training materials, manuscripts, publications, reference documents, Presentations in the form of Audio, Video files and dashboards, with password-protected access.
- **External**
  - ✓ Institutional repository of data with password-protected access
  - ✓ Institutional repository of on-line data collection tools
  - ✓ Data archives (as specified by funding agency, password-protected access)

8. UPDATING/ REVIEWING THE POLICY

This policy can be updated every 2 years or sooner if required by changes in policies/laws.

9. MANDATE FOR NEW PROJECTS

All the new Proposals by all PHFI/IIPH Departments/Divisions/Projects/ commencing with effect from February 2017, will incorporate a cost for utilizing PHFI CRDR Platform, as under:

Up-to 30 Lakhs	**No charge from the Project**
>Rs. 30 Lakhs to 1Crore	**Rs.25000/- Per Project**
>Rs. 1Crore to 5 Crore	**Rs.100000/- Per Project**
>Rs. 5 Crore to 10 Crore	**Rs. 150,000/- Per Project**
>Rs. 10 Crore and above	**Rs.200000/- Per Project**

In addition, the Closed Projects which are of high significance and important as a record for future Research purposes are under the preview of VP, Research & Policy to consider the cost for migration from the Central fund.

- ✓ The cost of Ongoing Project can be recovered from Contingency or any other Heads as per the Project Heads defined in the Proposal.
- ✓ It may be noted that the PHFI CRDR Platform will require continuous Migration Cost for new Studies in addition to AMC and therefore, the proposed cost recovery is recommended to meet the required Budget.
- ✓ To have full control on the quality of the collected Data, this Application will have Data Entry Module, EDC/Paper based along with Discrepancy tool for management of queries, SAE Reconciliation, External Data Load, Field specific reference Document loading facility, Mobile app, Dashboards and Report generation and Data Base lock facility.
- ✓ The Core PHFI CRDR team will develop the Survey forms using this Application and the PI Teams can perform all the DB related Project work using this Application. It is a 21 CFR Part 11 Compliant Application hence, can be used both for CTs' and other Projects.

10. RATIFICATION OF THE POLICY

This policy will come into effect after approval by the concerned policy review committee and any subsequent revisions as per the management decisions.

11. HANDLING OF BREACH OF THE POLICY AGREEMENT

Violation of this policy agreement in general and failure to comply with any of the above requirements in particular will be handled by the agreement signed between the two parties (i.e., the requestor and the PI) and other appropriate laws governing such activities.

12. LEGAL ISSUES PERTAINING TO DATA SHARING

The PHFI Internal Technical Review Panel and Core CRDR team will give appropriate consideration to legal issues in safeguarding the rights of participants, communities, organizations and intellectual property rights.

13. ETHICAL ISSUES PERTAINING TO DATA SHARING

The PHFI Internal Technical Review Panel and Core CRDR team will give appropriate consideration to ethical issues in safeguarding the rights of study participants. All the informed consent forms and data would be anonymized and made available after 3 years/ on completion of the projects, to others without personal identifiers and an ability to track down individual study participants.

14. FREQUENTLY ASKED QUESTIONS ON DATA SHARING

S. No	*Questions*	*Answers*
1	**Why should I share my final research data?**	Data sharing achieves many important goals for the scientific community, such as :- 1. Reinforcing open scientific inquiry 2. Encouraging diversity of analysis and opinion, 3. Promoting new research, testing of new or alternative hypotheses and methods of analysis 4. Supporting studies on data collection methods and measurement 5. Facilitating education of new researchers 6. Enabling the exploration of topics not envisioned by the initial investigators 7. Permitting the creation of new datasets by combining data from multiple sources.
2	**Who benefits from data sharing?**	Everyone benefits, including investigators, funding agencies, the scientific community, and, most importantly, the public. Data sharing provides more effective use of resources by avoiding unnecessary duplication of data collection. It also conserves research funds to support more investigators. The initial investigator benefits, because as the data are used and published more broadly, the initial investigator's reputation grows.
3	**How soon after data collection am I obliged to share the final data?**	At the earliest or after completion of project. **At the earliest** is suggested because it will help the PI to perform data validation, trend analysis to understand the project progress vis a vis - resource utilization, project progress, Milestone tracking, Budget tracking, Site performance opting for **Go-No GO decision** on an ongoing basis and also can perform fraud detection of data and can plan for addressing the data mismanagement. This way you can maintain integrity that is reliability and validity of data collection.
4.	**What do you mean by final research data?**	By "final research data", we mean recorded factual material commonly accepted in the scientific community as necessary to validate research findings. Final research data do not include laboratory notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens.
5.	**What do you mean by unique data?**	By "unique data" we mean data that cannot be readily replicated. Examples of studies producing unique data include: large surveys that are too expensive to replicate; studies of unique populations, such as centenarians; studies conducted at unique times, such as a natural disaster; studies of rare phenomena, such as rare metabolic diseases.

6	**Is data sharing widely accepted as a good practice?**	National scientific organizations have made a commitment to the sharing and archiving of data through their ethical codes (e.g., the American Sociological Association) or publication policies (e.g., the American Psychological Association). More than 15 years ago, the National Academy of Sciences described the benefits of sharing data. (See http://books.nap.edu/catalog/2033.html) For many years, the National Science Foundation (NSF) Economics Program has required data underlying an article arising from an NSF grant to be placed in a public archive. Similar expectations exist at the National Institute of Justice. Moreover, many scientific journals require that authors make available the data included in their publications. In the biological sciences, protein and DNA sequences are made available to researchers through data archives, such as GenBank. Since 1996, NIH has required data sharing in several areas, such as DNA sequences, mapping information, and crystallographic coordinates. For example: The PLoS Data Policy states that authors wishing to publish a paper in one of their journals must ensure that: Data that underpins their research findings is made available to others in an appropriate form. A Data Availability Statement is provided in the paper itself which outlines how data can be located…
7	**What kinds of data are candidates for sharing?**	Potentially all kinds of data are candidates for sharing, but unique data are especially important. Some biologic sciences already have data-sharing plans in place, such as genetic mapping. But other basic science data are also amenable to sharing. Data from human subjects (e.g., surveys, clinical studies) are also shared by protecting the identity and privacy of research participants.
8	**Can you give me some examples of data that have been shared?**	Examples of shared epidemiologic data include the Framingham Heart Study, the Honolulu Heart Program, the Atherosclerosis Risk in Communities, Epidemiology of Chronic Disease in the Oldest Old, and the Iowa 65+ Rural Health Study. Examples of shared data from clinical trials include the Asymptomatic Cardiac Ischemia Pilot, the Intermittent Positive Pressure Breathing Study, and the Safety and Efficacy Trial of Zidovudine for Asymptomatic HIV Infected Individuals. Examples of shared datasets from the basic sciences include a growing number of genome sequences and maps, as well as protein and nucleotide databases (see ENTREZ http://www.ncbi.nlm.NIH.gov/Database/index.html and other resources for molecular biology at the National Center for Biotechnology Information at http://www.ncbi.nlm.NIH.gov)
9	**Does data sharing pertain only to published data?**	No. Data-sharing plans should encompass all data from funded research that can be shared without compromising individual subjects' rights and privacy, regardless of whether the data have been used in a publication. Furthermore, data sharing prior to the publication of major results is discouraged in many instances (for a restricted period till the time all the major results are published), for example, when data are collected to

		provide a resource for the scientific community (as in the case of many large surveys.).
10	**Due to circumstances beyond my control (an earthquake!), I was unable to recontact a substantial portion of the sample. I was planning to put my data in an archive, but the resulting high rate of attrition makes the data minimally useful. Should I still archive the final dataset?**	Investigators need to find a balance between the value of the final data and the costs associated with archiving. If the data are of limited usefulness, then it is probably not worth the expense and effort of putting them in an archive. However, if the investigator has published results based on this dataset, then the dataset should be shared.
11	**Am I required to submit a data-sharing plan always?**	Yes. The specific nature of the data you will collect will determine whether or not you may share the final dataset. If the final data are not amenable to sharing, for example, if they are proprietary, then you need to explain this in your application.
12	**I don't want to share my data, which were generated under an PHFI grant. Can I be forced to do so?**	When the PI and the authorized institutional official sign the face page of a PHFI application, they are assuring compliance with policies and regulations of PHFI. PHFI expects grantees to follow these rules and to conduct the work described in the application. Thus, if an application describes a data sharing plan, PHFI expects that plan to be enacted.
13	**Will the data-sharing plan affect the priority score of my application?**	No. Reviewers will not factor the proposed data-sharing plan into the determination of scientific merit or priority score. They will assess the appropriateness and adequacy of the proposed data-sharing plan
14	**My research, which seeks support from both the public and private sectors, will involve proprietary data. How do I deal with the data-sharing issue in my application?**	PHFI recognizes that there may be circumstances where a co-funder has requested restrictions on data sharing as a condition of funding. These restrictions should be identified in the application and a proposal made about how data from the co-funded project will be shared. Should you believe that you are unable to share any of the data, your justification will be considered by PHFI review committee.
15	**I'm a busy investigator. I don't have time to process requests for my data. What should I do?**	Put data in the PHFI repository facility which provides a web site for data access and technical assistance for users with questions or problems. This may spare busy investigators time.
16	**Can I share data with colleagues under my own terms and conditions?**	No. Your data-sharing plans should indicate whether or not you will place any conditions on their use. Data should be made as widely and freely available as possible while safeguarding the confidentiality of the data and privacy of participants. You should not place limits on the questions or methods others might pursue. As per PHFI policy, PI would be given co-authorship in any research publication arising from the particular dataset.

17	**Should the data source be cited or acknowledged in papers that rely on shared data?**	It is appropriate to acknowledge the source of data upon which a manuscript is based. Many investigators include this information in the methods and/or reference sections of their manuscripts. Journals generally include an acknowledgement section, in which the authors can recognize people who helped them gain access to the data.
18	**Should I consider contributing my research data to PHFI data Repository?**	Definitely. PHFI repository will collect and distribute data. They understand what is needed to prepare data for wider distribution and documentation for users. They provide stable, reliable, and cost-effective means for distributing data. They also provide protections for the dataset and technical assistance for requestors.
19	**How to prepare data for sharing and archiving?**	Data to be shared should be in either of these formats: XLS,PDF, DOC,PPT,XPT,.CSV,.GIF,.AVI,.MP4,MP3, .WMP and .JPEG format
20	**How do I pay for preparing data for sharing and archiving?**	It takes time and money to prepare data for sharing. You can request funds for data archiving and sharing as part of your grant application for collecting the data. PHFI recommends that you consider procedures and costs for data sharing during the application process rather than after the data have been collected.
21	**Should I address data sharing in my application?**	Yes.
22	**What do I need to include in my application and where do I put the information about data sharing?**	Scientists submitting grant, cooperative, or contract applications should include a data-sharing plan, or provide justification for the absence of such a plan, in a brief paragraph. Additional information (after **Research Plan section**) on data sharing might be included in other sections of the application, as appropriate. For example, if you are producing a large dataset that will become an important resource for the scientific community, you probably want to mention this in the **significance section**. If you are requesting funds to prepare, document, and archive the data, you would want to include relevant information in the **budget and budget justification sections.** In the **Human Subjects section** of the application, you should discuss the potential risks to research participants posed by data sharing and steps you will take to address those risks.
23	**The informed consent form for my recently completed study states explicitly that only my research team will see the data provided and that we will not share the data. Am I now expected to share it?**	No, In preparing and submitting a data-sharing plan during the application process, investigators should avoid developing or relying on consent processes that promise research participants not to share data with other researchers. Such promises should not be made routinely or without adequate justification described in the data-sharing plan.
24	**Can institutions and investigators subject to**	Yes. PHFI recognizes that data sharing may be complicated or limited, in some cases, by institutional policies or local IRB/IEC

	the Federal Health Insurance Privacy and Portability Act (HIPAA) Privacy Rule share data in accord with the PHFI Data Sharing policy?	rules, as well as by local, state laws and regulations like the Privacy Rule. To protect the rights and privacy of people who participate in sponsored research, data intended for broader use should be free of identifiers that would permit linkages to individual research participants, and exclude variables that could lead to deductive disclosure of the identity of individual subjects. When data sharing is limited, applicants should explain such limitations in their data sharing plans.
25	**I collect data on sensitive and, sometimes, illegal behaviours. Are these data too sensitive to be shared?**	Not necessarily. The collection of sensitive data does not preclude sharing. For example, the National Center for Chronic Disease Prevention and Health Promotion at CDC operates the Youth Risk Behavior Surveillance System (YRBSS), available at https://www.cdc.gov/chronicdisease/index.htm , which provides data on six health risk behaviors among youth: unintentional injuries and violence, tobacco use, alcohol and other drug use, sexual behaviors, dietary behaviors, and physical activity. Similarly, data from the National Survey of Family Growth, which includes statistical data on family life, marriage and divorce, contraception, sexual experience, pregnancy, and infertility, can be obtained from the National Center for Health Statistics - US. Sensitive data can be shared so long as appropriate privacy safeguards are in place. Investigators must determine if and how the rights and privacy of the subjects can be protected.
26	**Can data from a clinical trial be shared?**	It depends. Participants' privacy must be protected in accord with all applicable laws and regulations. Clinical trial datasets are frequently rich in items that could potentially identify individual subjects. For example, many early phase trials use small samples, which make it difficult to protect the privacy of the participants. Researchers who are planning clinical trials and intend to share the resulting data should think carefully about the study design, the informed consent documents, and the structure of the resulting data prior to the initiation of the study. Should ensure de-identification of the personal information of the subjects before sharing the data. There are many precedents for sharing of clinical trial data. For example, data from a number of clinical trials supported by the National Heart, Lung, and Blood Institute (NHLBI) are available for research use (See https://www.nhlbi.nih.gov/health-topics/publications-and-resources). The National Institute of Allergy and Infectious Diseases (NIAID) also lists their clinical trials datasets that they have made available through the National Technical Information Service (NTIS) for public use (See https://www.niaid.nih.gov/clinical-trials).
27	**Is data on DNA and protein sequences archived?**	Yes. For example, GenBank (https://www.ncbi.nlm.nih.gov/genome) and Entrez (http://www.ncbi.nlm.NIH.gov/Entrez/) archive gene sequencing data. The sharing of materials, data, and software in a timely manner has been an essential element in the rapid

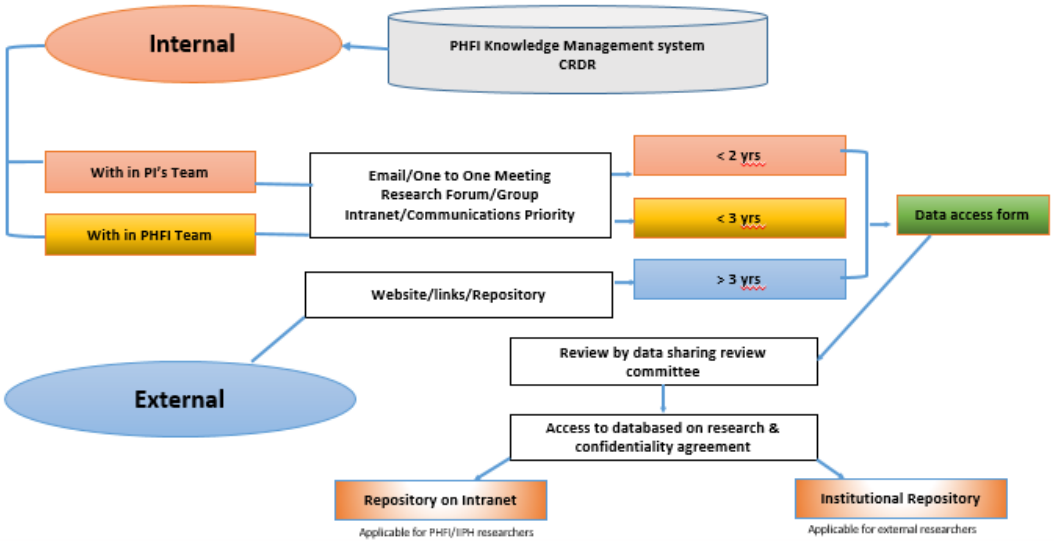
		progress that has been made in the genetic analysis of mammalian genomes.
28	**I did not request support for sharing data in my application, which was funded. Can I charge PHFI for the costs associated with sharing the data?**	No, you just have to upload the data (as per prescribed format) in the free FTP folder access to which given to you by the PHFI. It will take hardly few minutes to upload large volume of data.
29	**I am working on a select pathogen and cannot share the data for reasons of national security. Is this an acceptable reason for not sharing?**	Yes, Sharing can be curtailed but only by due process involving RMC and not at the discretion of the investigator.
30	**If I am required to submit a revised data-sharing plan, what do I need to do?**	As is the case with PIs who submit any additional or revised application material, your revised data-sharing plan must be signed by PHFI authorized signatory and by you
31	**I want to request a dataset from a recent publication. How do I do this?**	You should check the publication to see if reference is made to PHFI repository, where the data might be available. If no such information is provided, you may wish to send a letter to the PI to see if the data are available for sharing, and where you might be able to get the data and associated documentation.
32	**Who all are major funding agencies of PHFI?**	Major funding agencies such as Welcome Trust, National Institute of Health, NHLB, MRC, ICMR, Department of Biotechnology etc.

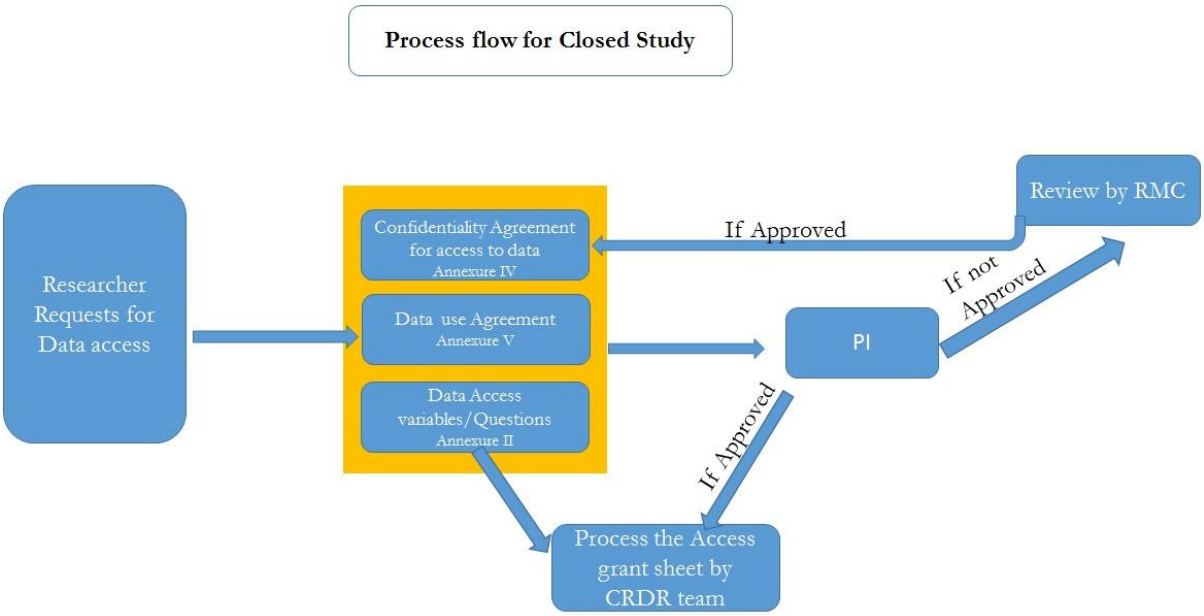
15. REFERENCES

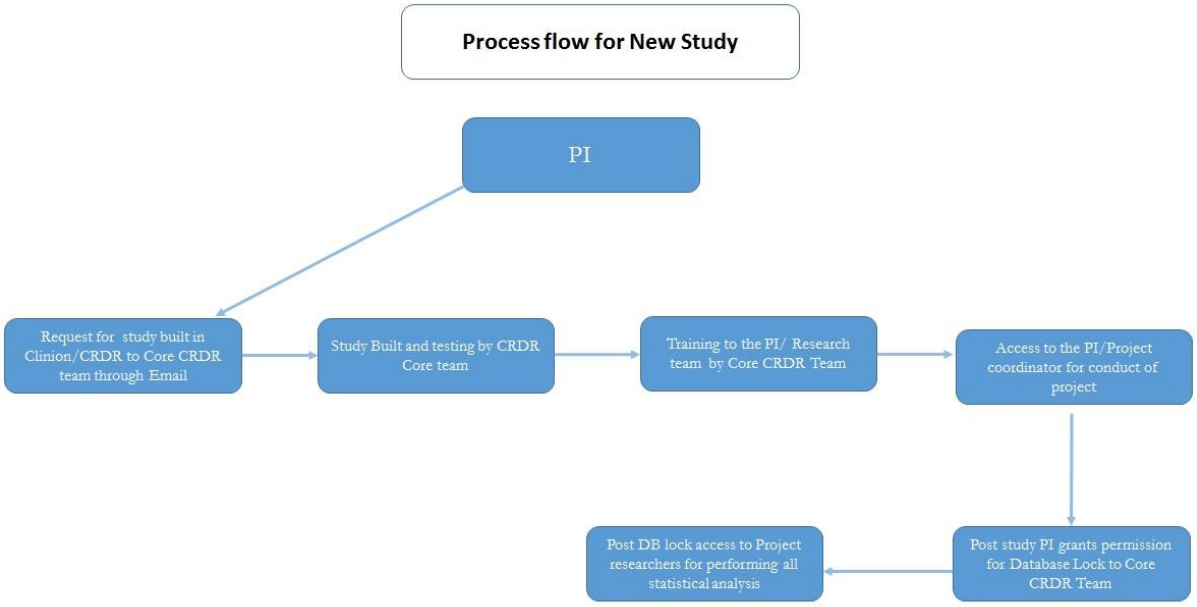
1. http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm
(last accessed on 19/11/2020)
2. http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-andepidemiology/WTDV030690.htm
(last accessed on 18/11/2020)
3. http://grants.nih.gov/grants/policy/data_sharing/
(last accessed on 19/11/2020)
4. http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/
(last accessed on 18/11/2020)
5. http://icmr.nic.in/ethical_guidelines.pdf
(last accessed on 20/11/2020)
6. http://www.nhlbi.nih.gov/research/funding/human-subjects/data-sharing.htm
(last accessed on 6/01/2017)
7. http://www.who.int/bulletin/volumes/88/6/09-074393/en
(last accessed on 18/11/2020)
8. http://books.nap.edu/catalog/2033.html
(last accessed on 19/11/2020)
9. http://www.lshtm.ac.uk/eph/ncde/groups/epigenetics/funders/index.html
(last accessed on 16/11/2020)
10. http://en.m.wikipedia.org/wiki/Video_file_format
(last accessed on 16/11/2020)

16. ANNEXURES

Annexure I: Data Access Flow Diagram






Process flow for New Study
PI
Request for study built in Clinion/CRDR to Core CRDR team through Email
Study Built and testing by CRDR Core team
Training to the PI/ Research team by Core CRDR Team
Access to the PI/Project coordinator for conduct of project
Post study PI grants permission for Database Lock to Core CRDR Team
Post DB lock access to Project researchers for performing all statistical analysis

Annexure II: Data Access Variables/Questions

PART A: Basic Details required for Data Access

S.no	Activity	Yes/No	Remarks(if any)
1	Applicant details (name, contact information, institution) Attach detailed CV _____ _____		
2	Proposed Project Title: _____ _____ _____		
3	Aims and objectives of proposed project (max 150 words): _____ _____ _____ _____		
4	Justification for use of the PHFI data and data requirements, and analysis to be conducted(max 250 words):		
5	Is the work used for a degree program?		
6	Will the work be funded and if so, by whom? _____ _____		
7	Are biological samples required, and if so, what are eligibility criteria for participant samples, which tests are to be conducted and how much sample will be required? (Please refer to the Biorepository guideline of PHFI/ discuss with the concerned Research Officer/PI) _____ _____		
8	Projected timeline of project _____		
9	Are there any ethical issues associated with this work and will new ethical approvals be required? _____ _____		

PART B: Confidentiality agreement Form for Data Access and Usage– PHFI

Instructions: **Access to PHFI data (i.e., the right to examine PHFI data) is strictly limited under PHFI guidelines. Persons seeking access to data in the custody of PHFI or IIPH must complete this application, sign and submit it to the custodian of the data.** **Attach additional pages as necessary. Approval will be mailed to Applicant at the address shown above.**
1. Applicant requires access to PHFI data to engage in the following demographic, epidemiological or other similar studies related to health:
2. The specific purpose for which Applicant will use PHFI data and the data files to be accessed (e.g. type(s) of cancer, patient characteristics, diagnosis years, geographical areas) and other relevant information are:
3. Applicant's qualifications to engage in these activities are as follows:
4. In consideration for PHFI Data Custodian's approval of this application, Applicant represents, warrants, and agrees as follows: For purposes of this confidentiality agreement, "PHFI data" means all information relating to cases of study as per point no 1 mentioned above collected at any time by PHFI or any other individual or institution. PHFI data also means all documents, files or other records, regardless of format or medium, containing PHFI data. **PHFI data sharing policy** contains various provisions relating to use, access, disclosure, and publication of PHFI data. These provisions may be different from the laws, regulations or policies applicable to other data used by Applicant.
Applicant represents and warrants that: Applicant has reviewed the "**Data Sharing policy** from the portal www.CRDR.PHFI.org "for Access to and Disclosure of Confidential Data from the PHFI repository and the terms and conditions of this confidentiality agreement . (**Yes/No**)
Notwithstanding any other provision of this confidentiality agreement, PHFI Data Custodian shall have no obligation to grant Applicant access to PHFI data unless and until his or her application is approved.
By my signature I declare as follows:
I have read the foregoing agreement. By signing below I make the agreements and representations contained therein. I understand that these are material representations of fact upon which reliance was placed when this transaction was entered into.
Signature Dated
Name & Designation
APPROVAL BY PHFI DATA CUSTODIAN:
Signature Dated
Name & Designation

Annexure III: Summary Of Funder's Requirements For Data Sharing

Funder	*Requirements*
Action Medical Research (AMR)	**Data Outputs Covered:** All research outputs.
	Data Plan Requirement: It does not require a data management/sharing plan.
	Funding arrangement for data management and sharing: None stated
	Documentation requirements: Clear and accurate records should be retained on procedures followed and approvals granted during research process, including interim and final results.
	Retention requirements: Data should be retained for a minimum of 10 years.
	Data Standards: None stated
	Publication Requirements: Researchers are expected to make data available at the same time as research results are published. Publication may be delayed for a 'reasonable period', pending, intellectual property protection, ethics approvals, or consents being obtained.
	Data Sharing Rights: None stated
	Designated data centre: None stated
	Monitoring: None stated
	Institutional Requirements: Institutions should publish documents detailing the standards that they encourage to enable good research practice as well as procedures for investigation research misconduct allegations and Institutions should provide training for new researchers and students on data management, record keeping, data protection, Intellectual property management, and other research techniques.
Biotechnology and Biosciences Research Council (BBSRC)	**Data Outputs Covered:** All research outputs.
	Data Plan Requirement: A 'Statement on Data Sharing' plan is required for all applications
	Funding arrangement for data management and sharing: Funding for data management and sharing may be requested as part of Full economic cost for the project.
	Documentation requirements: Documentation should be created that describe the data's provenance and enable its content to be understood. Researchers are expected to make use of current guidance on best practice..
	Retention requirements: Data should be retained for a minimum of 10 years after completion of research project.
	Data Standards: Researchers expected to store data in accessible formats using established standards
Biotechnology and Biosciences Research Council (BBSRC)	**Publication Requirements:** Researchers are expected to make data available at the same time as main findings of research are published or in line with best practice in field. Explicit reasons must of research are published or in line with best practice in field. Explicit reasons must be provided in Data Sharing Plan if data publication is not possible or appropriate. The BBSRC reserves the right to implement a more prescriptive approach to data sharing for specific research initiatives.
	Data Sharing Rights: IPR remains with investigator and institution. Data should be made available with as few restrictions as possible.
	Designated data centre: No designated data Centre. Deposition in an appropriate data archive or data enclave is encouraged. Alternatively, direct

	sharing by investigator or institution, or mixed mode sharing may be appropriate.
	M**onitoring:** Compliance monitored through Final Report assessment procedure. Data sharing of previously produced research is taken into account when assessing future proposals.
	Institutional Requirements: Institutions receiving BBSRC funding should possess guidelines setting out responsibilities and procedures for maintaining data.
Bill & Melinda Gates Foundation	**Data Outputs Covered:** Final, annotated quantitative and qualitative datasets and accompanying information such as metadata, codebooks, data dictionaries, and questionnaires. Data may arise as a primary output of a grant, or as a product of other activities, such as program budget..
	Data Plan Requirement: A 'Data Access Plan' must be prepared for grants over $500,000.
	Funding arrangement for data management and sharing: Funding for data management and sharing may be requested as part of the project budget.
	Documentation requirements: Documentation should be created that describe the data's provenance and enable its content to be understood. Researchers are expected to make use of current guidance and information on best practice.
	Retention requirements: None stated
	Data Standards: Researchers should use standards appropriate to the subject domain.
	Publication Requirements: Data should be made available in a 'timely manner'. The grantee may specify a time period for exclusive use of data (e.g. 12-18 months), after which the data will be made available for others. Alternatively, the data may be released with a Data Use Agreement, indicating that the user may analyses the data, but is not allowed to publish on topics related to the grantee's area of research. 3rd party proprietary data is not expected to be made available. Data should be shared with as few restrictions as possible; Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.
	Data Sharing Rights: None stated
	Designated data centre: Deposition in a public access data archive or data enclave is encouraged. Alternatively, direct sharing by investigator or institution, or mixed mode sharing may be appropriate
	Monitoring: None stated
	Institutional Requirements: None stated
Wellcome Trust	**Data Outputs Covered:** All data outputs, including records of interim results and final research outcomes.
	Data Plan Requirement: A 'Data Management and Sharing Plan' is required for applications to the Trust's biomedical sciences and medical humanities funding streams if the project will produce data outputs that have value to the research community in the short and/or long-term.
	Funding arrangement for data management and sharing: Data management and sharing costs may be incorporated into the project budget.
	Documentation requirements: Documentation should be created that describe the data's provenance and enable its content to be understood. Records should be created and maintained that describe the procedures

	followed, approvals granted, and results (interim and final) found during the research process.
	Retention requirements: Data should be maintained in a secure environment for a minimum of 10 years.
	Data Standards: Data should be maintained in a secure environment for a minimum of 10 years
	Publication Requirements: Data should, if possible, be made available at the same time as research papers on the findings are published. Data creators have the right to delay or limit data sharing to safeguard research participants or to address intellectual property issues. Any Such restrictions should, however, be minimized and a clear statement provided on the reason for delays should be provided in the data management and sharing plan.
	Data Sharing Rights: None Stated
	Designated data centre: An appropriate subject or institutional repository should be used. A list of possible options may be found at < http://www.wellcome.ac.uk/About-us/Policy/Spotlightissues/Data-sharing/Guidance-for-researchers/WTX060360.htm >
	Monitoring: Grant holders are encouraged to liaise with the funder to address data management issues. All grant holdess are asked to report back on their approach for disseminating their research as part of their end of grant report.
	Institutional Requirements: Data generated in the course of research should be kept securely in paper or electronic format, as appropriate. Institutions should possess documentation establishing responsibilities and procedures for the storage and disposal of data and samples (including compliance with requirements of any ethics committee).
WHO - World Health Organization – TDR	**Data Outputs Covered**: A definition of data outputs is not provided, but may cover all research data produced using project funding.
	Data Plan Requirement: The Data management requirements vary between each funding call. Applicants are asked to describe the research lifecycle within their proposal, describing the technical and practical activities associated with data cleaning, monitoring, and verification
	Funding arrangement for data management and sharing: Not stated
	Documentation requirements: No specific requirements are stated. However, the project will need to document their work in accordance with Good Clinical Practice.
	Retention requirements: Data should be retained for a minimum of 10 years.
	Data Standards: None stated
	Publication Requirements: None stated
	Data Sharing Rights: None stated
	Designated data centre: Selected datasets are published on the Global Health Observatory Data Repository (http://apps.who.int/ghodata/)
	Monitoring: None stated
	Institutional Requirements: None stated
World Cancer Research Fund (WCRF)	**Data Outputs Covered**: Definition of data outputs is not provided, but may cover all research data produced using project funding
	Data Plan Requirement: WCRF does not require researchers to submit data management or sharing plans in grant applications.

	Funding arrangement for data management and sharing: Funding is not provided for indirect costs, such as institutional overheads associated with data management
	Documentation requirements: None stated.
	Retention requirements: In accordance with the UK Department of Health Research Governance framework, data collected in the course of research must be retained for an 'appropriate period' to allow further analysis by the original and/or other research teams (subject to consent)
	Data Standards: Research Governance must conform with institutional policies on topic and/or UK Department of Health Research Governance framework
	Publication Requirements: In accordance with the UK Department of Health Research Governance framework, research findings and data relevant to findings must be "made available for critical review through accepted scientific and professional channels" and "made accessible" to those participating in the study, as well as other who could benefit from them. No time limit is stated.
	Data Sharing Rights: IPR remains with the host institution. The WCRF reserve the right to make copies of data for national members
	Designated data centre: Research findings are added to the WCRF/AICR Continuous Update Project (CUP) database http://www.wcrf.org/cancer_research/cup/)
	Monitoring: Funder requires report on project progress, but does not state an explicit requirements for the underlying dataset
	Institutional Requirements: Institution must possess policies on research governance
National Health Service Technology Assessment (NHS HTA)	**Data Outputs Covered**: All outputs produced by the project
	Data Plan Requirement: Applicants are encouraged to describe their approach to data sharing in the project proposal, plan and monthly reports. However, the application does not provide a formal template for the data management plan.
	Funding arrangement for data management and sharing: Unstated
	Documentation requirements: Documentation should conform to Department of Health (DoH) Research Governance Framework and Medical Research Council guidelines for Good Research Practice
	Retention requirements: In conformance with MRC guidelines, primary research data must be retained by the producing research institution for a minimum of 10 years after project completion. Research records relating to clinical or public health studies should be retained for a Minimum of 20 years.
	Data Standards: Projects must follow the Department of Health (DoH) Research Governance Framework and Medical Research Council guidelines for Good Research Practice.
	Publication Requirements: No time limit is stated. The NHS NTA request that they are provided with a copy of research outputs to be published at the time of submission or at least 28 before the publication date
	Data Sharing Rights: IPR remains with the host institution. Monographs published by projects through HTA are covered by Crown copyright
	Designated data centre: The HTA programme is working with the National Cancer Research Unit (NCRI) Informatics Initiative to formalize and co-ordinate data sharing efforts. Data may be deposited with a data enclave or data archive, or shared through an institutional website or through posting CDs. Investigators may wish to use a data-sharing agreement to define

	criteria for data access, including allowed use and confidentiality standards that should be met.
	Monitoring: None stated
	Institutional Requirements: None stated
Natural Environment Research Council (NERC)	**Data Outputs Covered**: All research data
	Data Plan Requirement: A 'Data Management Plan' is required for all projects that will produce data outputs
	Funding arrangement for data management and sharing: Not stated
	Documentation requirements: Data should be catalogued according to the NERC profile of ISO19115:2003
	Retention requirements: Data and research materials should be preserved and made available for a minimum of 10 years after completion of the research. NERC indicate that research outputs of major importance may need to be retained for 20 years or longer**.**
	Data Standards: Specified in consultation with relevant data centre
	Publication Requirements: Project may define an embargo period of up to two years from end of data collection to enable researchers to analyse and publish findings. Potential data users Encouraged to contact data creator to negotiate for earlier use.
	Data Sharing Rights: IPR remains with host institution. NERC data centres are allocated a non-exclusive license to manage and publish research data
	Designated data centre: Deposit with authorized NERC data centre (BADC, NGDC, NEODC, BODC, PDC, EIDC, UKSSDC, ADS). Catalogue record published through NERC Data Catalogue Service
	Monitoring: Ongoing monitoring. Grant holders who do not meet these requirements risk having award payments withheld or become ineligible for future funding
	Institutional Requirements: None stated
Medical Research Council (MRC)	**Data Outputs Covered**: All research data
	Data Plan Requirement: A 'Data Management Plan' is required for all projects that will produce data outputs
	Funding arrangement for data management and sharing: Not stated
	Documentation requirements: Projects are encouraged to use the Data Documentation Initiative (DDI) metadata standard
	Retention requirements: Primary research data must be retained in their original form by the research institution that generated them for a minimum of ten years from completion of the Project. Research records relating to clinical or public health studies should be retained.
	Data Standards: The MRC recommend that grant holders refer to UK Data Archive, Inter-university Consortium for Political and Social Research (ICPSR), Inter-university Consortium for Political and Social Research (ICPSR), Australian National Data Service and National Statistics Code of Practice for information on appropriate standards.
	Publication Requirements: Data should be made available in a 'timely and responsible manner'. A limited period of exclusive use of data for primary research is considered reasonable .

	Data Sharing Rights: The study must publish a Data-sharing policy establishing appropriate license (e.g. Open Data Commons) for use of anonymized data
	Designated data centre: Research data should be submitted to the MRC Research Data Gateway.
	Monitoring: None Stated
	Institutional Requirements: None stated
NIH	**Data Outputs Covered**: All final research data
	Data Plan Requirement: A 'Data Management Plan' is required for all projects where funding is more than $500,000
	Funding arrangement for data management and sharing: Can request funds for data sharing and achieving.
	Documentation requirements: Good Documentation practice to be adhered.
	Retention requirements: 3 years following the end of a grant .
	Data Standards: Not stated
	Publication Requirements: Final project report .
	Data Sharing Requirement: Sharing of data no later than the acceptance for publication of the main findings from the final dataset… in timely manner.
	Designated data centre: No specific archive mentioned.
	Monitoring: Final project report
	Institutional Requirements: None stated
DBT/DST/ICMR	**Data Outputs Covered**: None stated
	Data Plan Requirement: None stated
	Funding arrangement for data management and sharing: None stated
	Documentation requirements: Good Documentation practice to be adhered.
	Retention requirements: None stated
	Data Standards: Not stated
	Publication Requirements: Full text of the research paper along with metadata and supplementary materials should be deposited to either institutional Repository or Centralized repository. .
	Data Sharing Requirement: Sharing of data no later than the acceptance for publication of the main findings from the final dataset… in timely manner.
	Designated data centre: None stated
	Monitoring: Project report/On Publication of research results
	Institutional Requirements: None stated

- For genome-wide association studies funded by NIH the grantees are expected to provide a plan for submission of GWAS data to the NIH-designated GWAS data repository, or provide an appropriate explanation why submission to the repository is not possible. http://gds.nih.gov/

According to Welcome Trust policy for data sharing, "***In some cases it may not be appropriate for researchers to share their data. However, if the research meets the criteria for requiring a data management and sharing plan but you are intending not to share your data, the reasons for this must be clearly justified***" (http://www.wellcome.ac.uk/Aboutus/Policy/Spotlightissues/Data-sharing/Guidance-for-researchers/WTX060360.htm)

Annexure IV: Publisher's Requirement

Publisher	Requirements
LANCET	***Research for Life/HINARI*** All countries in the HINARI programme can get access to all of the content on The Lancet website for free or at very low cost through The Lancet journals' arrangements with HINARI. This initiative has the full support of Elsevier. In addition, developing countries can access our content via Science Direct through Elsevier's partnership with HINARI. ***Access to TheLancet.com via geographical-based IP addresses*** All content published in The Lancet publications is freely available to readers residing in developing countries, as defined by the UNDP Human Development Index, via the recognition of geographical-based IP addresses (Geo-IP). ***Sponsored articles*** Articles reporting research that has been funded by agencies that we have a sponsorship agreement with are made available free of charge on The Lancet.com. Ref: www.thelancet.com/information-forauthors
PLOS	Access to research results, immediately and without restriction, has always been at the heart of PLOS' mission and the wider Open Access movement. However, without similar access to the data underlying the findings, the article can be of limited use. For this reason, PLOS has always required that authors make their data available to other academic researchers who wish to replicate, reanalyze, or build upon the findings published in PLOS journals. In an effort to increase access to this data, PLOS have revised the data-sharing policy for all PLOS journals: authors must make all data publicly available, without restriction, immediately upon publication of the article. Beginning March 3rd, 2014, all authors who submit to a PLOS journal will be asked to provide a Data Availability Statement, describing where and how others can access each dataset that underlies the findings. http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/
Bio Med Central (BMC)	Submission of a manuscript to a BioMed Central journal implies that readily reproducible materials described in the manuscript, including all relevant raw data, will be freely available to any scientist wishing to use them for non-commercial purposes. Well established and widely supported databases exist for certain types of data such as nucleic acid sequences, protein sequences, and atomic coordinates; information on which can be found below and in journal instructions for authors and 'about' pages. An increasing number of research funding agencies also now support data sharing in the life sciences. BMC Open Data policy aims to clarify the legal (copyright) status of data published in its open access journals and to maximize the potential for reuse of published science, such as data and text mining. For society to gain the full benefit from scientific data, it needs to be available such that it can be reused, scrutinized and

	built upon with the minimum of barriers – in accordance with the Panton Principles for Open Data in Science. This means enabling reuse of data without needing special permission from its original creators by waiving copyright and related rights in published data. To achieve this, unless otherwise stated in an individual article's license, data included in BioMed Central's published open access articles are distributed under the Creative Commons CC0 1.0 Public Domain Dedication waiver. Anyone reusing data published in BioMed Central journals must, wherever possible, cite the source(s) of the data in a derivative work, although this is not a legal requirement. The Creative Commons CC0 waiver applies to data included in the article, its reference list(s) and its additional files. We have described in detail the case for using Creative Commons CC0 for data in a 2012article in BMC Research Notes. Ref: http://www.biomedcentral.com/about/opendata
BMJ	The BMJ requires sharing of individual patient data for all clinical trials. This means that trials will be considered for publication only if the authors agree to make the relevant anonymised patient level data available on reasonable request. The BMJ is the first general medical journal to require data sharing for all trials, extending its initial policy on sharing data for trials of drugs or devices, which took effect in January 2013, says Elizabeth Loder, The BMJ's acting head of research. Ref: http://www.bmj.com/company/wp-content/uploads/2015/07/data-sharing-policy-20151.pdf The BMJ asks authors of original research articles to state in their manuscripts whether they are making available any additional unpublished data. These may comprise raw unprocessed data as well as protocols, analyses, statistical codes, images, and ideas (http://resources.bmj.com/bmj/authors/types-of-article/research). BMJ ask this largely because they are keen to maximize the usefulness and usage of data and promote transparency, but also because many research funders now encourage or even mandate data sharing.1 Many BMJ articles' authors simply say "no additional data available," but a growing repository of positive data sharing statements range from "an audit trail of the forest plots and related data is available at www.wolfson.qmul.ac.uk/bptria "2 to "a full list of participants' quotes and explanations offered by the authors to illustrate each of the four themes are available on request from the corresponding author at rachaelm@health.usyd.ed.au . 3 BMJ are delighted that authors have been so willing to share data.

Annexure V: Format Data use Agreement – PHFI

Data Use Agreement

This data use agreement (the "Agreement") is by and between Public Health Foundation of India ("PHFI"), a society registered under the Indian Societies Registration Act, 1860, having its registered office located at Public Health Foundation of India, ISID Campus, 4th Institutional Area, Vasant Kunj, New Delhi- 110017 and Correspondence address : Sector 44, Plot 47, Institutional Area, Gurgaon 122002, India, and the < Requestor> and is effective as of < day of Month, 2020 (the "Effective Date">.

WHEREAS, PHFI maintains certain information that User wishes to use and/or disclose for research purposes:

NOW, THEREFORE, the parties, in consideration of the mutual promises and obligations set forth herein, the sufficiency of which is hereby acknowledged, and intending to be legally bound, agree as follows:

1. PHFI shall provide User with access to certain data (XXXX) in accordance with the terms and conditions of this Agreement.
2. The following individuals (the "Authorized Parties") are authorized to use the Data Set or any part of it on behalf of User and agree to abide by the terms of this Agreement.

Name: _____ Signature: _____

Name: _____ Signature: _____

Name: _____ Signature: _____

3. PI will define the terms and conditions as per the grant and the institutional policy.

   Authorship:

   The PI should be given Co- Authorship in any research publication arising from the particular dataset.

   _____

   _____

   _____

   _____

   Rights:

   IPR remains with Investigator and Institution.

   _____

   _____

   _____

   _____

   Sharing with others:

Data shall not be transferred to other users in any form, must adhere to permissions/approvals. Shall not use the data to promote any illegal activities or presenting datasets in a misleading manner.

_____

_____

_____

_____

Financial Implications:

Funding for data sharing may be requested as part of full economic cost for projects.

_____

_____

_____

_____

Future grant applications, if any:

_____

_____

_____

_____

4. User and each Authorized Party agrees as follows:

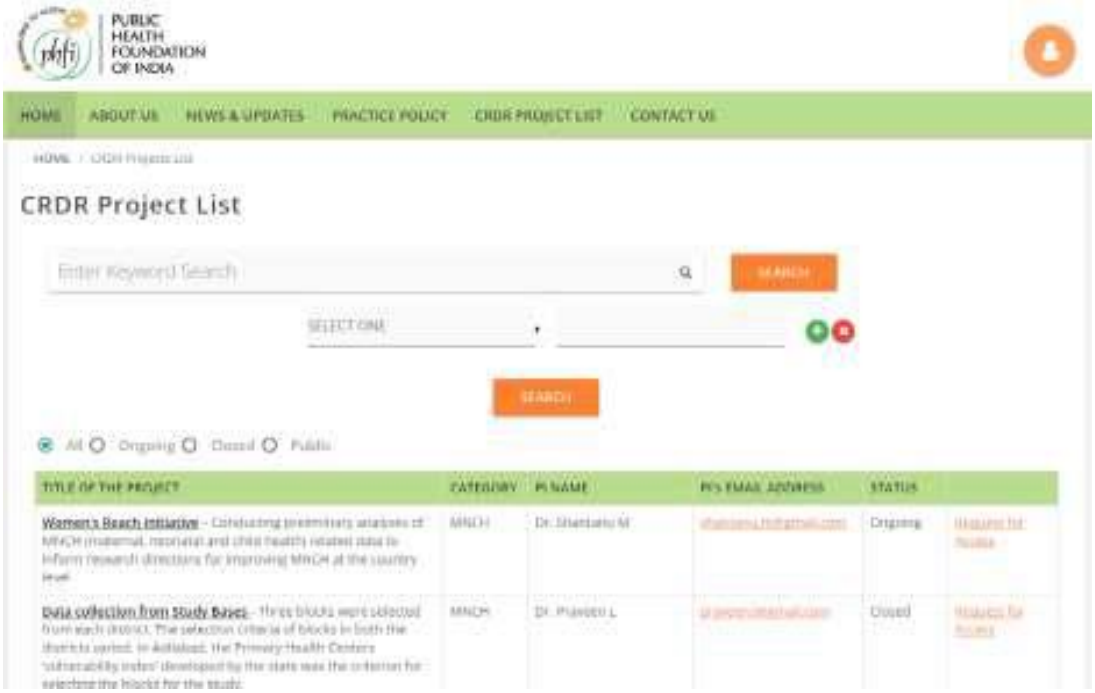
- ✓ Any publications arising from the data would be reviewed by PHFI for authorship based on contributions prior to being published.
- ✓ Not to use or further disclose the < Data Set> or any information contained therein other than as permitted by this Agreement or required by applicable law.
- ✓ To use appropriate safeguards to prevent use or disclosure of the information other than as provided for by this Agreement.
- ✓ To report to PHFI any use or disclosure of the Baseline Data Set or any part of it not provided for by this Agreement of which User or any Authorized Party becomes aware.
- ✓ To ensure that any User employees, agents, or subcontractors, to whom User or an Authorized Party provides the Baseline Data Set or any part of it to agree to the same restrictions and conditions that apply to the User and Authorized Parties under this Agreement.
- ✓ Not to use the information contained in the Baseline Data Set to identify the individuals whose information is contained in the Baseline Data Set, nor to contact them under any circumstances.
- ✓ To destroy or return the Baseline Data Set at the completion of the purpose identified above.

5. In the event PHFI becomes aware of any use of the Baseline Data Set or any part of it that is not authorized under this Agreement or required by applicable law, PHFI may (i) terminate this Agreement upon notice; (ii) disqualify (in whole or in part) the User and/or any Authorized Parties from receiving any data in the future; and/or (iii) report the inappropriate use or disclosure to the proper authority.

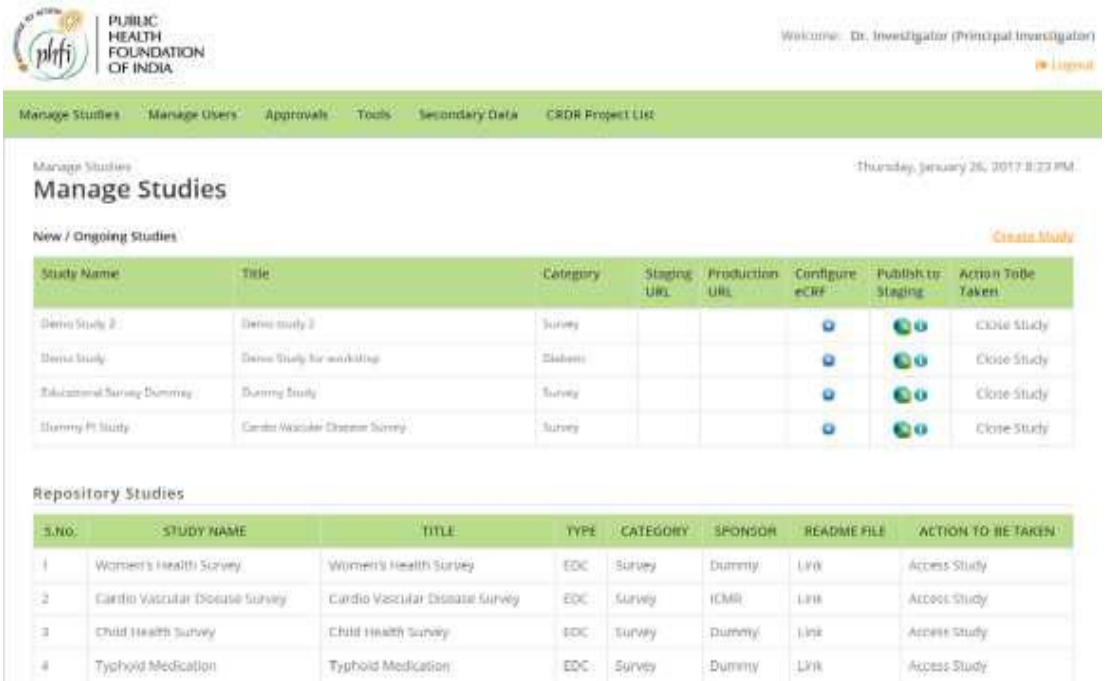
   WHEREFORE, the parties, through their authorized representatives, hereby accept and agree to the terms and conditions of this Agreement.

PHFI	USER
Authorized Signature:	**Authorized Signature:**
Name (Printed):	**Name (Printed):**
_____	_____
Title:	**Title:**
_____	_____
Institute:	**Institute:**
PUBLIC HEALTH FOUNDATION OF INDIA	_____
Date:	**Date:**
_____	_____

Annexure VI: Repository Management Information Layer Application Features:

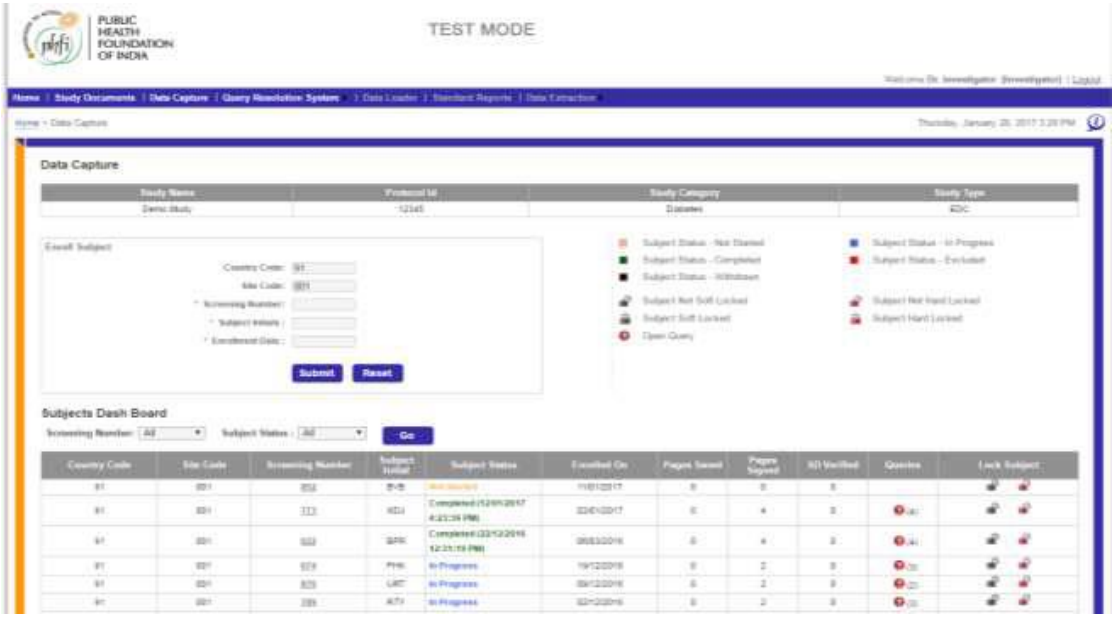


- ✓ **Master Data Management**: The application will allow the Principal investigators to manage their projects, create new projects, close/restrict or move them to repository.

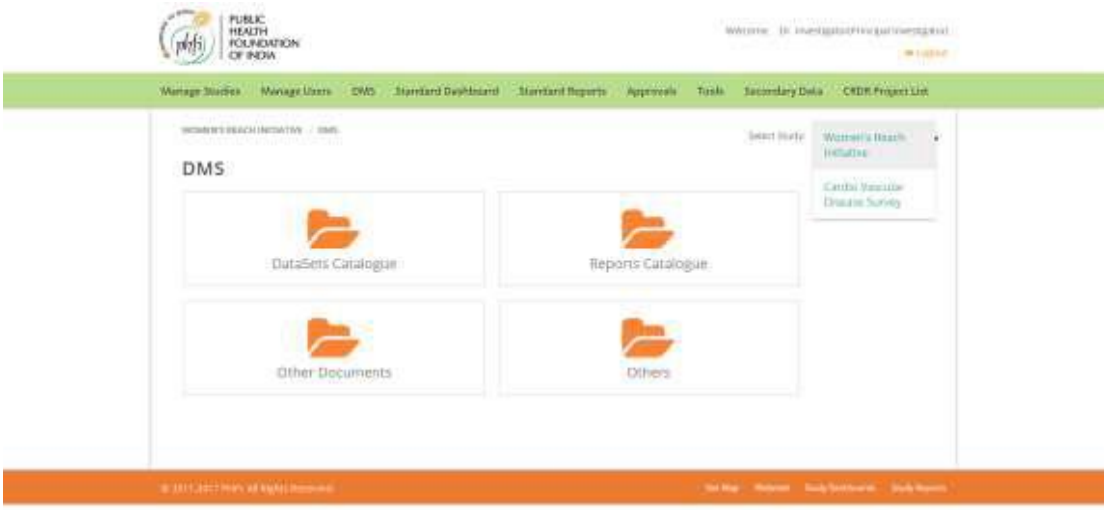


- ✓ **Data Migration**: The repository studies will be initially migrated by the Quad One team, into the application. After the migration of 155 studies (as per the scope), a tool will be integrated into the system which will enable the users to migrate future closed studies into the application.
- ✓ **Integration**: The application will have third party integration of tools which are freely available. These tools will be provided as links which can be downloaded by the end users on their systems. These tools could be used for conversion of file formats or to upload large sized files into the system, etc.

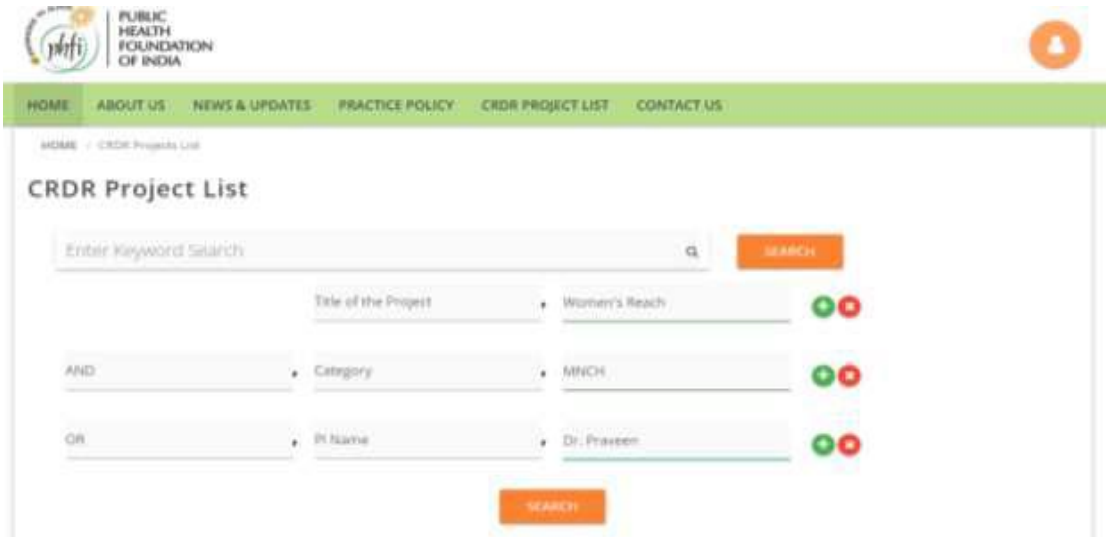
- ✓ **Electronic Data Capture**: The system will link to the Clinion application for Electronic Data capture. Clinion provides a user friendly interface for enrolling subjects and performing data entry.



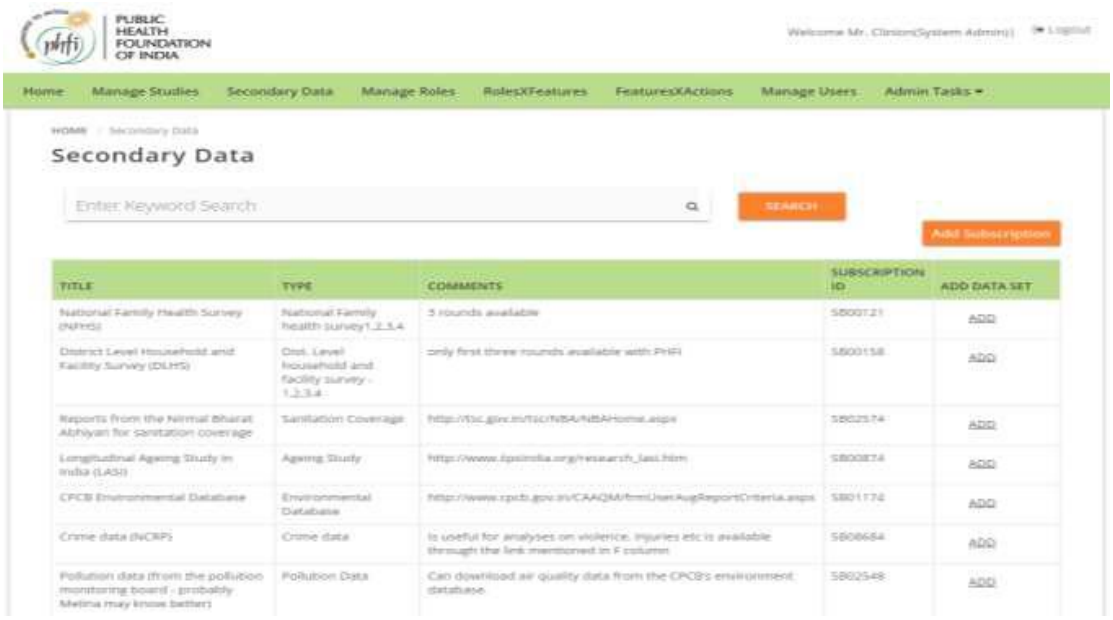
- ✓ **Document Management System**: The application will provide a platform for the users to upload their project specific, analysis reports into the system. The DMS will have data sets catalogue, reports catalogue and other documents related to specific projects.



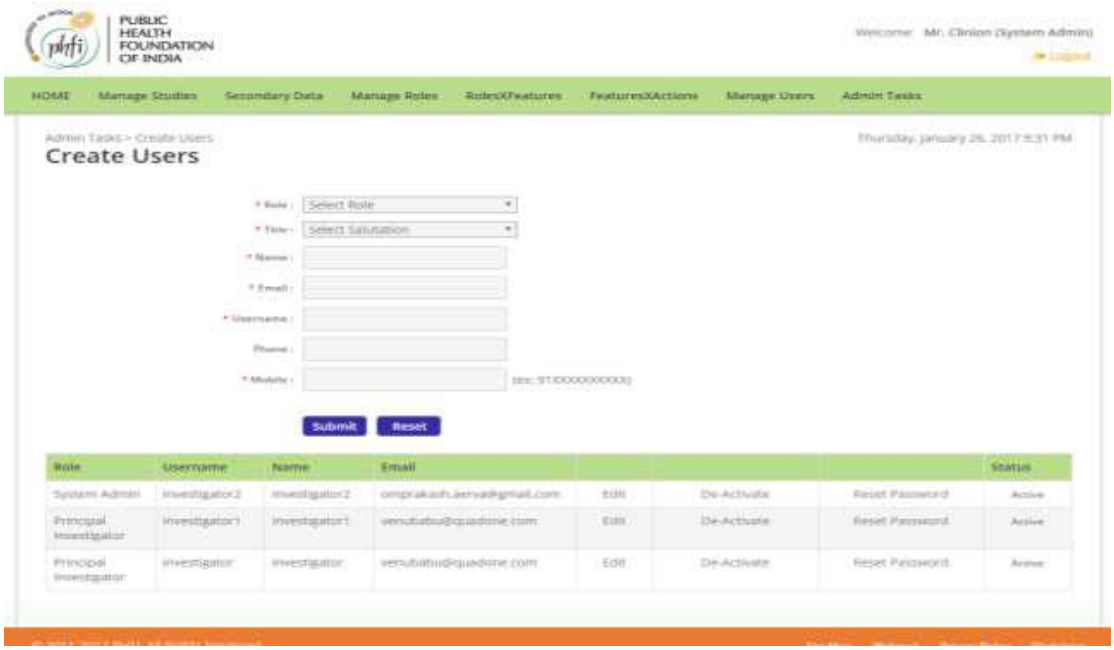
- ✓ **Search**: The application will provide 'keyword search' and 'Advanced search' options, for optimizing the search results of the users. These search options will be provided in the 'CRDR project' list and wherever applicable. The Advanced Search options will have categories, 'Title of the project', 'Category', 'PI name', 'Project Status', 'Date Submitted', 'Last 30 days', 'Last 60 days'. These categories can be added with 'AND' and 'OR' options.



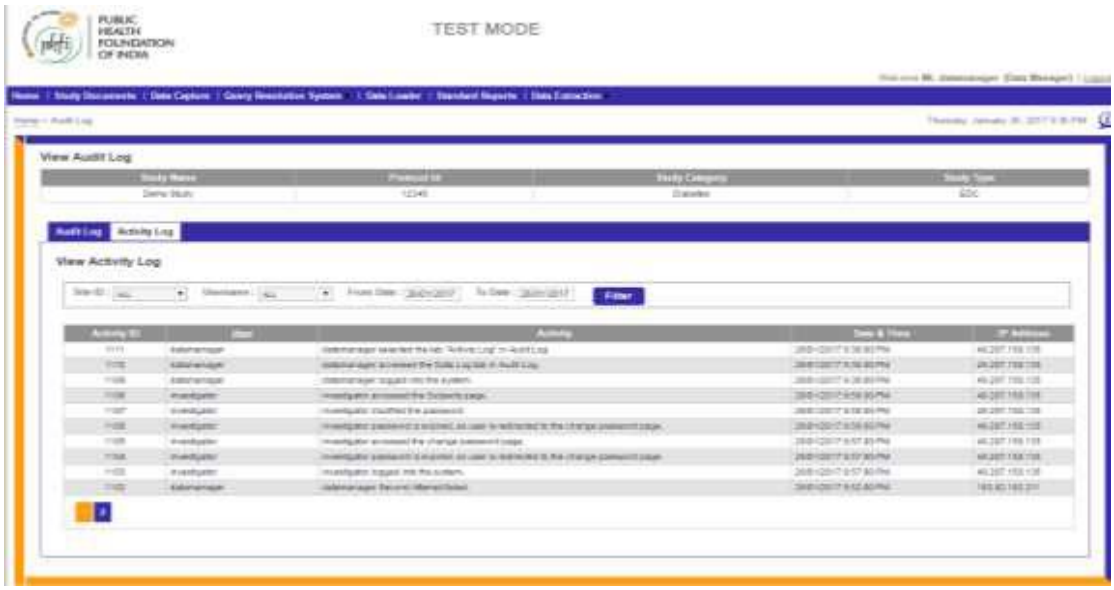
- ✓ **Ongoing Secondary Data**: Secondary data can be uploaded into the application, by the SYS ADMIN of CRDR. Access to paid secondary data can be provided to the users, after the Principal Investigator's approval. The SYS ADMIN of CRDR can also upload data sets to the already existing secondary data.



- ✓ **User Management**: The application will allow the SYS ADMIN of CRDR and the Principal Investigators to create users and assign them to projects. User management includes creation of role, user, assigning projects, assigning features and permissions to each role for using the Application.

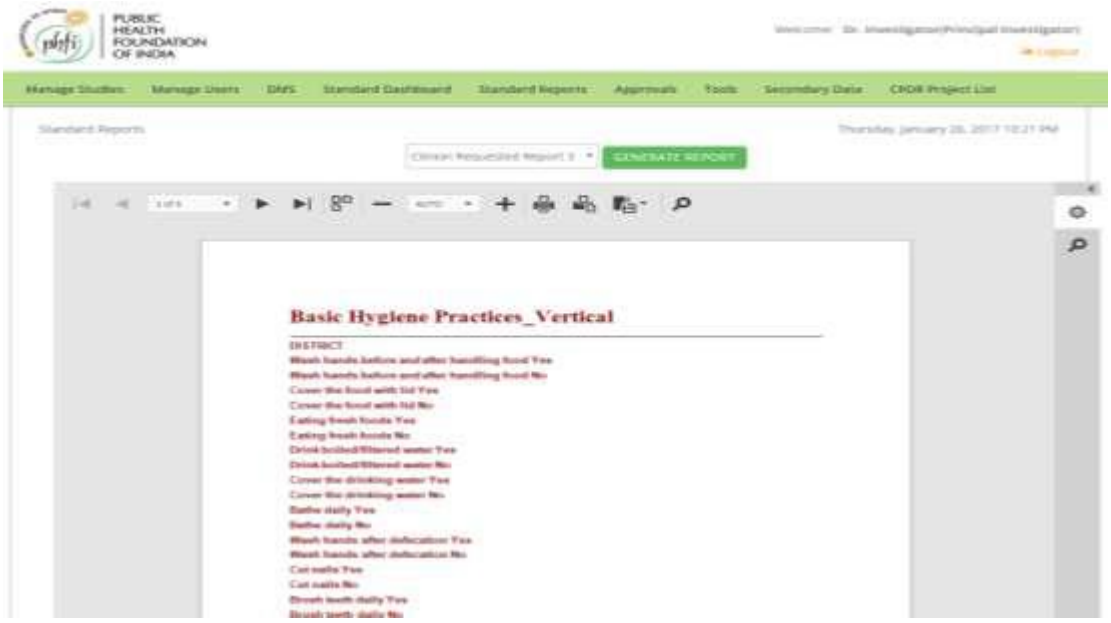


- ✓ **Security**: The application will be highly secure. Users will get locked out of the system, on making three incorrect password attempts. Moreover, the application will also maintain an Activity log' which will keep a track of 'click wise' activities performed by the users, during their access. The Activity log will provide the 'username, activity performed, date and time stamps and the IP addresses' of the users.

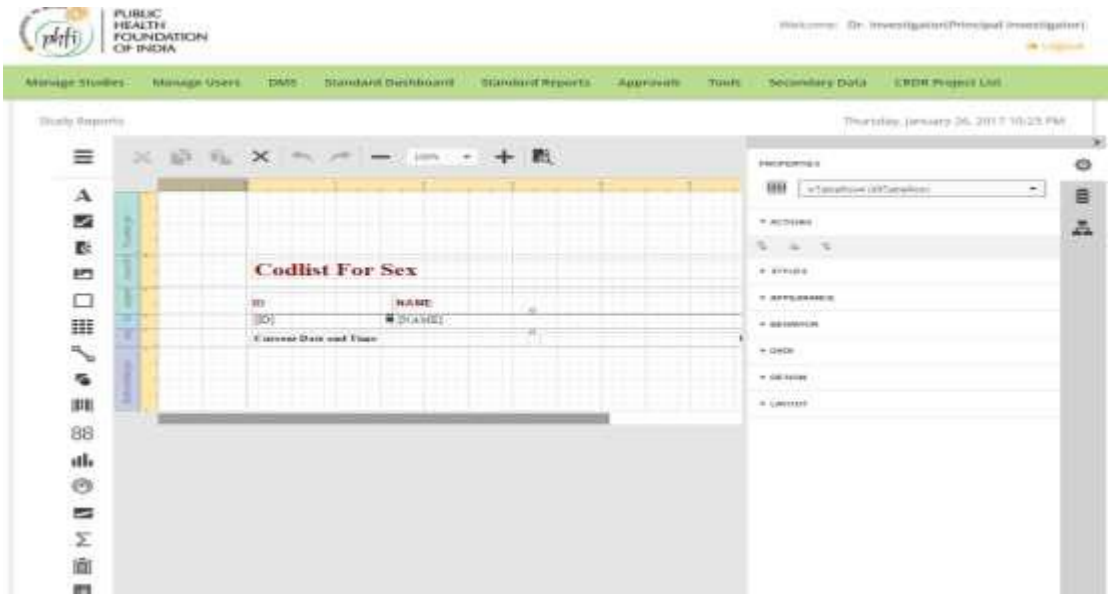


- ✓ **Audit**: The application will also maintain a record of the 'Audit log'. The data entered into the application during data capture will be recorded in the 'Audit log' section of the application. The system will also maintain a record of the file download and upload history in the DMS.

- ✓ **Reports**: The application will have Standard Reports and Study specific reports. The standard reports will be default for all the studies, however, the study specific reports can be created by the users, within the application by using the available tools. The users can also download the data in their systems, perform analysis and upload the data back into the DMS.

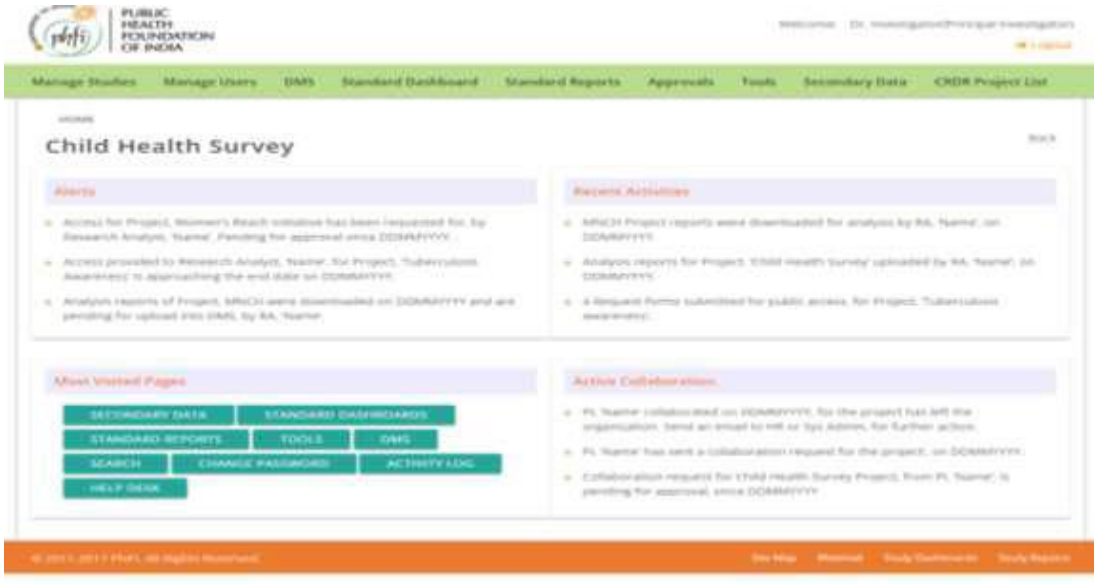


Reports: Standard Reports

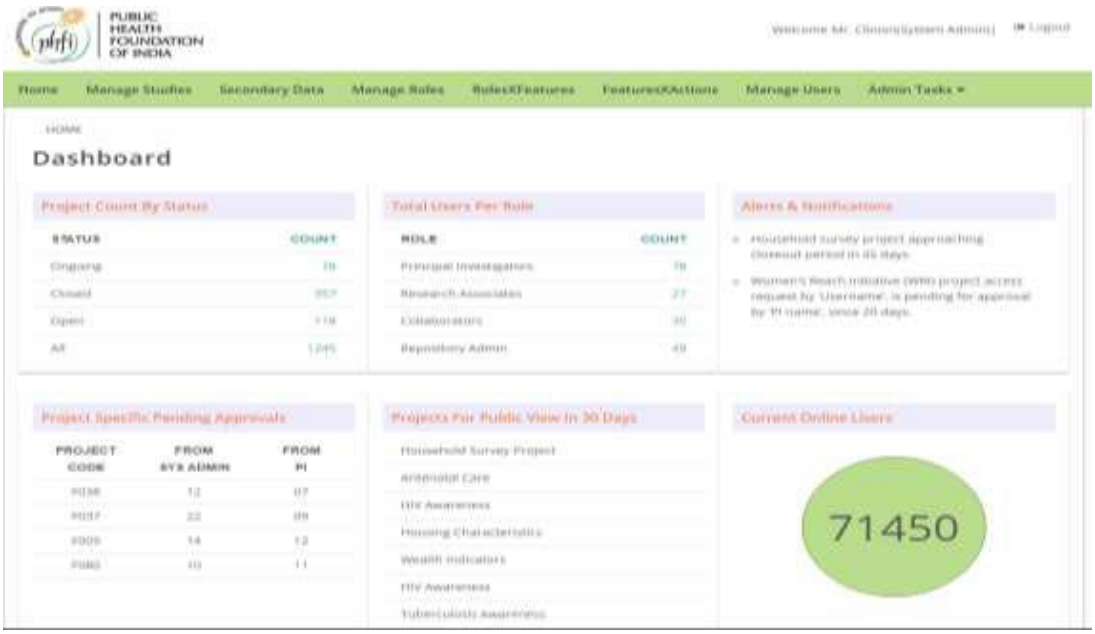


Reports: Study Specific Reports With Tools

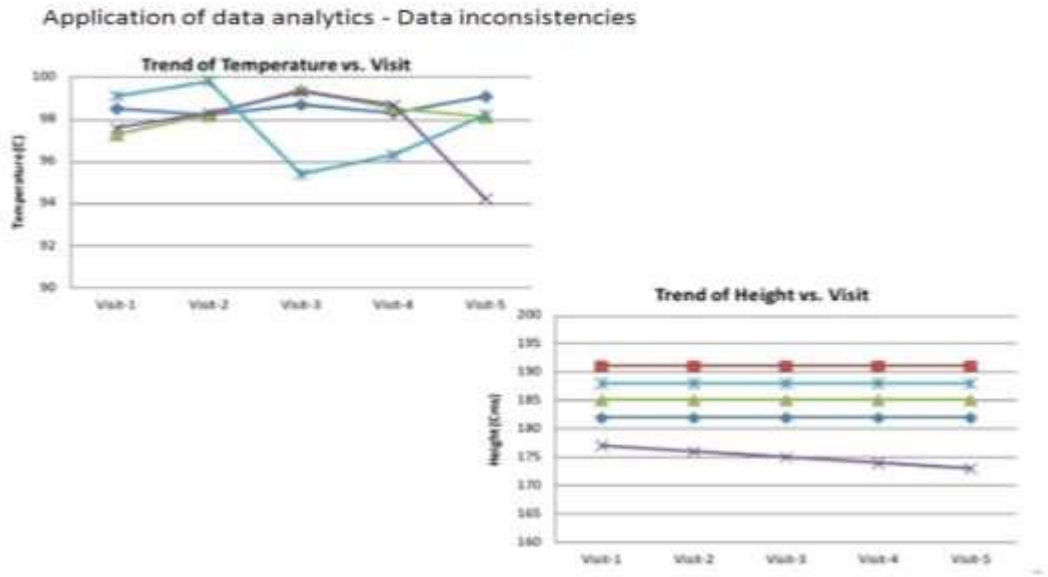
- ✓ **Dashboard**: The application will have dashboards for all the users, based on their access levels and permissions for each project. The application will also have project specific dashboards.



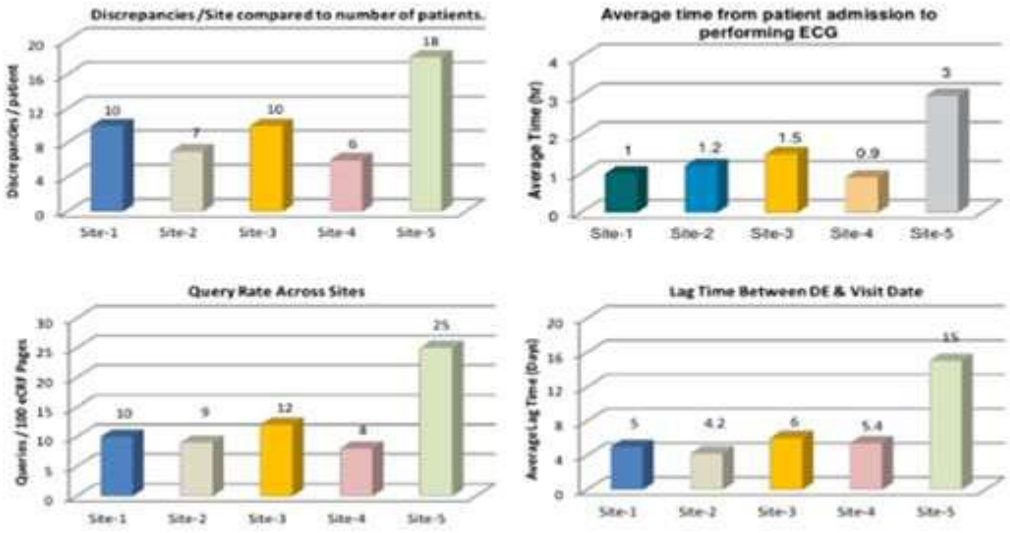
Dashboards: Project Specific Dashboard (PI)



Dashboards: SYS ADMIN of CRDR's Dashboard



Snapshots of the Data Dashboard



Application of Data Analytics – Data Trends and Outliers

End Of Document