2020

DRC DATA SHARING POLICY

Dabat Research Center

Table of Contents

Preface

This document incorporates details of data sharing policy, issues related to publications and authorship at the Dabat Research Center (DRC). The principles of data sharing are widely recognized and underpin many international activities. Publicly-funded research data are a public good, produced in the public interest and it should be openly available to the maximum extent possible. Data generated from DRC is available with as few restrictions as possible in a timely and responsible manner to the scientific community for subsequent research. In order to give appropriate credit to each author of a paper, the individual contributions of authors to the manuscript should be specified. An 'author' is generally considered to be someone who has made substantive intellectual contributions to a published study.

List of Abbreviations

DRC   Dabat Research Center

DUA   Data Use Agreement

HDSS   Health and Demographic Surveillance System

MoU   Memorandum of Understanding

PI   Principal Investigator

UoG   University of Gondar

1. Introduction

The generation, dissemination and utilization of demographic and health related evidence through census, vital events registration, population-based surveys and surveillance systems, are fundamental for evidence-based teaching, policy formulation and decision, program planning and practice. Countries generate these types of information through different combinations of the above methods, depending on the type and level of the required information and the resources available to them. Most developed countries conduct vital events registration, census and national population-based surveys and surveillance systems on regular basis [1]. Unfortunately, such evidences are generally limited in developing countries mainly due to resource constraints. In Ethiopia, one of the major sources of information is the Population and Housing Census which is conducted roughly every ten years [2]. The other sources of information in the country included health institutions-based records [3] and national surveys [4].

Although these nationwide surveys were used as sources of information, they were conducted infrequently. Moreover, there has been no continuous systematically organized registration of vital events. In order to fill the information gap to some extent, the University of Gondar (UoG) has established the Dabat Research Center (DRC) also known as the Dabat Health and Demographic Surveillance System (Dabat HDSS) site in 1996 at the Dabat district which is located approximately 821 km Northwest of Addis Ababa and 75 kms north of Gondar town. The research center generates longitudinal information on vital events and estimates population-level cause of deaths using verbal autopsy method from 70,000 population living in 4 urban and 9 rural kebeles. The DRC site covers three geo-climatic zones (cold, temperate and hot) which may reflect some pictures of the Amhara region related to health and demography.

In the near future, the UoG foresees to establish two more HDSS sites at Gondar town and Gorgora District, considering urban population dynamicity, ecology diversity, transboundary population movement, and water source (e.g. Lake, Sea, and Dam) induced public health problems. The sites will be a platform for multidisciplinary research (health, social science, and agriculture) to generate valid and reliable evidence that can be transformed by policymakers and planners into action.

1.1. Rationale of Data Sharing Policy

Publicly-funded research data are a public good, produced in the public interest and publicly-funded research data should be openly available to the maximum extent possible [5, 6]. Data-sharing is an important way to increase the ability of researchers, scientists and policy-makers to analyze and translate data into meaningful information and knowledge [7]. Sharing data strengthens open scientific investigation, encourages diversity thinking and collaboration, promotes new ideas and research, makes possible new hypotheses testing and methods of analysis, discourages duplication of effort in data collection and supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not covered by the initial investigators, encourages accountability and transparency enabling researchers to validate one another's findings, and permits the creation of new datasets when data from multiple sources are combined allowing comparisons of research findings across regional and national boundaries [5, 7].

Despite huge benefits gained from data sharing, there has been no clear guideline on how to implement it in DRC. Consequently, DRC has taken the initiative of developing and regularly reviewing a data sharing policy document that will guide the process of data sharing. This policy document sets out the roles of each party to the agreement in relation to the data shared and their responsibilities therein. This agreement concerns individuals and household survey datasets and the HDSS longitudinal dataset containing anonymized information. The data generated under DRC will be made available with as few restrictions as possible in a timely and responsible manner while safeguarding the privacy of participants, and protecting confidential and proprietary data. DRC will share data to individuals, individual groups, or institutions for efficient utilization of the data stored at its research site.

1.2. Purpose of the Data Sharing Policy

The purpose of this data sharing policy document is to guide the effective, efficient, and appropriate/ethical utilization of data generated by DRC through the harmonization of national, international and collaborating partners' requirements and guidelines.

2. Types of DRC Data

DRC collects three forms of datasets:

1. *Survey Data:* This includes the individual and household survey datasets that will be shared and managed centrally in the research center at the end of the data collection.

2. *Qualitative Data:* This includes data from the focus group discussions, in-depth interviews and transcripts created from these. This data will be anonymized.

3. *HDSS Data:* This includes the key indictors from rural and urban HDSS longitudinal datasets that will be kept at the research center and will be shared in the form of a template for purposes of pooled analysis.

3. Types of Data Access

DRC considers the following types of data access for the purpose of this policy:

3.1.1.   Open Access Data

Data will be freely available for anyone who is interested to use. Getting permission from DRC is not necessary for the utilization of such data which are identified to be open for anyone. Such data can be released either through a public use dataset or can be requested. Open access data includes but not limited to the following:

- Data necessary to calculate sample sizes
- Data necessary for describing the socio-demographic characteristics of the study population e.g., age and sex distribution of the study population
- Data necessary for calculating fertility, mortality, migration, and marital status change rates
- Data necessary for analyzing the causes of death

3.1.2.   Restricted Access Data

DRC shall decide the type of data that will be shared as restricted access. Agreement shall be entered between DRC and the prospective data users with regard to the type of data that will be shared. A registration by a prospective data user is required on the research center data repository with the following minimum information:

a. Full name,

b. Email address,

c. Institutional affiliation,

d. Country,

e. Category of user (e.g., student, researcher, etc.),

f. A statement of purpose for which the data will be used,

g. An ethically secured proposal, and

h. An agreement to the conditions of use.

3.1.3. Closed Access Data

Data will not be shared with anyone for different reasons like the sensitive/confidential nature of the data. This applies to highly sensitive or individually identifiable data. Such data are normally available to prospective users only through controlled-on-site access and/or in collaboration with the research center.

3.1.4. Access that Needs Primary Data Collection

Researchers are welcome to collect primary data within the DRC demographic surveillance area (DRC data collection sites). In such case, the following applies:

- A memorandum of understanding or a data sharing agreement and use shall be signed between the researchers and DRC.

- The researchers shall cover all the costs associated with the data collection, management, and analysis. In addition, the researchers will be charged 20% of the data collection budget for administrative costs.

- The collected data shall be the property of DRC, except in a case where the ownership is explicitly specified in the data sharing agreement and use or in the memorandum o understanding.

- In the case where the owner is DRC, the collected and cleaned data shall be submitted to DRC before any publication for archival for further research outside the research objectives of the researchers.

4. Principles of Data Sharing

The DRC data sharing principles are set in line with the data sharing principles published by the World Data System Scientific Committee considering the data policies of a number of national and international initiatives [8]. In addition, Data Sharing Principles in Developing Countries have been considered [6]. The data sharing principles of DRC are as follows.

Accountability

Sharing data requires trust and accountability between parties. Hence, the data sharing policy shall enable trust and accountability.

- DRC shall be accountable for the data warehouse update and management
- DRC shall be accountable for assuring site level data quality, timely availability of data
- DRC shall be accountable for proper decision based on the data sharing policy

Confidentiality

Confidentiality ensures that shared data are anonymized or protected from unauthorized use, access and disclosure to the maximum extent possible to maintain the privacy and confidentiality of individuals' information. The data sharing policy shall specify that all shared data shall be strongly protected to avoid disclosure of personal data. DRC shall maintain confidentiality of data based on the ethical principles.

Data Quality

To generate evidences and make the best possible decisions, data shared by DRC must be of the highest possible quality. In public health, the most desired characteristics of data quality are completeness, consistency, timeliness, and accuracy [9-11]. DRC shall ensure that shared are complete, consistent, accurate, and available to users timely and have not been altered or corrupted in an unauthorized manner. DRC shall maintain data quality at its data warehouses.

Efficiency

DRC believes that shared data results in increased efficiencies in terms of reducing duplication of effort in capturing and acquiring data.

5. Guidelines for Data Access

Authorization for DRC data users will be granted upon decisions of a team led by the Director of DRC to grant authorization. This section guides how individuals and institutions can access DRC data for research purpose.

5.1.1. Data Access to Collaborative Partners

Modalities for data sharing and publication among partners should be established in a memorandum of understanding (MoU) before submission of research applications to donors or before implementation of research projects to avoid misunderstandings later. The following guidelines should govern the formation of research MoUs regarding data access and related issues:

1. Agreement on access to data shall apply to the data collected through the joint project. If there is a need to extend access to other related data collected by the Center, an agreement on this should be captured in the same or a separate MoU.

2. Access to data will be limited to partners that are mentioned in the proposal. Partners should not share/use the data with non-authorized persons. Unless specifically negotiated and authorized, partners may not use data without involvement of DRC staff as co-authors on scientific and other publications written using the data.

3. Partners shall fill out a Data Sharing and Use Agreement form and commit to abide by the agreement for data use specified on the form. Users "co-own" data generated by the joint projects and the primary purpose for filling the form is to keep track of who is using which data and for which purpose.

4. Access to data for partners will remain valid for two years after the end of the project to allow further analysis of data for scientific publication. This period may be extended if there is clearly defined outstanding work.

5.1.2. Data Access to External Users

- DRC may grant permission to individuals or institutions (governmental and non-governmental) to access data for research purpose depending on the type of data requested.

- Data may be shared to the public domain two years after the release of analytical datasets. The 24 months embargo is meant to enable investigators and other staff working on the project to finalize defined papers addressing core study objectives and allow other interested staff to use the data.

- External users can apply for access to the data only using the online application form and must abide by the guideline outlined on the form.

- External users may use publicly available DRC data for research. In this case, no requirement for involvement of DRC staff shall be placed on the use of such data. However, all ensuing publications must acknowledge DRC as the source of the data, and copies of the publication must be sent to DRC.

- External users seeking to use data that are not yet in the public domain may be given permission to use such data. In such case, DRC staff shall be involved in the research project as co-author (partner) and the data access to the external users shall be agreed upon an MOU guiding the partnership.

5.1.3. Data Access to Graduate Students

As part of its capacity building program, DRC encourages graduate students to use its data (particularly data that have been made publicly available) for writing their thesis and dissertations. Students seeking to use data that are not yet in the public domain may be given permission to use such data if they are supervised by collaborative partners and the data access to the students is agreed upon in the MOU guiding the partnership.

DRC may request that one of its staff members to be appointed to serve on the student's dissertation committee in instances in which the student requires access to non-publicly available data sets. In such cases, DRC expects the student to carry out his/her preliminary and other data

analyses at the Center. During this time, the student will function as a Research Intern and will be expected to contribute to other aspects of research at DRC, including data processing, analysis or management of fieldwork.

Students who have authorized access to data are bound by the guidelines and restrictions specified in the Data Sharing and Use Agreement form. For instance, the student shall not use the data for any other purpose other than the dissertation (unless specifically authorized to do so) and shall not share the data with any other party, including their professors (without written permission from research center). Students who intend to publish their work after completing their dissertation should seek special permission from Research center.

6. Data Sharing Strategies

This section provides information about the methods of data sharing and documentation or metadata.

6.1.1. Methods of Data Sharing

The method for data sharing that DRC selects is likely to depend on several factors, including the sensitivity of the data, the size and complexity of the data requested, and the volume of anticipated data sharing requests. The sharing can be done through secured electronic communication (e.g., email attachments), portable storage medium, or transferring to a data archive facility. DRC will need to determine which method of data sharing is best for a particular data sharing.

6.1.2. Documentation

Irrespective of the mechanism used to share data, DRC shall prepare documentation for the dataset to be shared. The purpose of the documentation is to ensure that researchers can use the dataset so as to prevent confusion, misuse, and misinterpretation. The documentation shall provide information about the sampling techniques and procedure, data collection methods and procedures, data dictionary, and definitions of variables. The precise content of documentation will vary by scientific area, study design, the type of data collected, and characteristics of the dataset.

7. Data Sharing Agreement

The DRC data sharing agreement shall be a formal contract that clearly documents what data are being shared and how the data can be used. The agreement is made between DRC and prospective data user. The agreement shall ensure that the data will not be misused. It shall also prevent miscommunication between DRC and the data recipient by establishing the parameters of data use. Each prospective data recipient shall complete a data sharing agreement with DRC.

There shall be four main components to the data sharing agreement: defining the data to be shared, securing the shared data, complying with any and all legal requirements on the data, and specifying the conditions under which data may be shared with external entities other than the data sharing participants.

The following items shall be included in the DRC data sharing agreement:

- Name of the data user
- Address of the data user
- Institutional affiliation of the data user
- Support letter from the institution where the data user is working
- Dataset to be given to the user
- Purpose for which the data will be used
- Timeframe the shared data will be used
- Whether it is possible to transfer the data to third party
- Mechanisms of maintaining confidentiality and security of the data
- Methods of data-sharing
- Authorship, if necessary
- Financial costs of data-sharing

8. Data Use Agreement

A Data Use Agreement (DUA) is an agreement used for the transfer of data that is subject to some restriction on its use [12]. A DUA addresses important issues such as limitations on use of the data, obligations to safeguard the data, liability for harm arising from the use of the data, publication, and privacy rights that are associated with transfers of confidential or protected data. A DUA is necessary when sharing data that are not de-identified in a manner that was not explicitly covered in a consent form. De-identified datasets may NOT contain any of the following but not limited to Names, Address, location, and any other unique identifying number, character, or codes.

The following statements shall be considered when using DRC data:

- Data originating from the research center may not be analyzed or reported without the explicit permission of DRC.
- Data and other material provided by DRC will not be redistributed or sold to other individuals, institutions or organizations without a written agreement with DRC.
- No attempt shall be made to re-identify respondents, and there shall be no use of the identity of any person or establishment discovered inadvertently. Any such discovery will be reported immediately to the research center.
- No attempt shall be made to produce links between datasets provided by DRC or between DRC data and other datasets that could identify individuals.
- Any books, articles, conference papers, thesis, dissertations, reports or other publications employing data obtained from the research center must cite the source, in line with the citation requirement provided with the dataset.
- An electronic copy of all publications based on the requested data will be sent to the research center.
- The original collector of the data, the research, and the relevant funding agencies bear no responsibility for the data use or interpretation or inferences based upon it.

In all the agreements, DRC will:

- Warrant that the necessary ethical and institutional approvals are in place for the sharing of data within the Project.
- Warrant that agreement for the sharing of the Data has been granted by the other project Principal Investigator(s) (PIs).
- Provide initial data structures for creation of a standardized dataset from the routine surveillance with selected variables of interest to the Project.
- Maintain ownership and responsibility for the upkeep of the Project Data provided.
- Supply and maintain on the repository an acknowledgement of the original data providers and a citation for use in publications.

This agreement stipulates that all parties agree to:

1. Ensure Data are preserved until all parties acting legally in its place, deem that these Data are to be destroyed or no longer required to be stored in the source.
2. Ensure Data will be stored in a secured manner with appropriate safety and routine maintenance.
3. Ensure that a suitable mechanism is provided to enable members of the Project to engage actively with the Data for the purposes outlined.
4. Ensure that access to Data is controlled through secure, authenticated mechanisms to prevent any unauthorized access.
5. Manage the accreditation status of individuals accessing the Data.
6. Provide an audit trail for any changes introduced to the Data over time.
7. Make available to the end-user a standard form of words, provided by DRC, for the purpose of acknowledgement of the DRC.
8. Ensure study teams' participation in design, analysis and manuscript writing arising from all projects that use the data accrued

9. Authorship Policy

Authorship provides appropriate credit for an individual's contributions to a research work and carries accountability. DRC believes that authorship credit shall be given to those who have substantial contribution and participation to a research work, and shall fairly and truly reflect actual contributions. Authorship determination through fair and equitable manner helps to maintain upright reputation of the University of Gondar in general and DRC in particular.

The purpose of this policy is to establish clear guidelines related to authorship of scholarly publications. This policy applies to all researchers who are using DRC data for scholarly publications. The scholarly publications include, but not limited to, books, articles, abstracts, presentation in scientific conferences, workshops, and meetings, and gran applications. The following statements apply in all scholarly publications using DRC data:

- Four criteria must be fulfilled for a researcher to qualify for authorship:
    1. Substantial scholarly contributions to conception and design of a research project, OR significant collection of data that needs significant intellectual input, OR analysis and interpretation of data for the creation of the research output;
    2. Drafting or revising the research output OR contributing critically important intellectual content to the article;
    3. Final approval of the version to be published; and
    4. Agreement to be listed as an author. That is agreement to be accountable for all aspects of the work, including, but not limited to, accuracy, integrity, or appropriateness the research work.
- A researcher who meets the above outlined authorship criteria must not be included or excluded as an author without his/her prior permission.
- One author shall take the primary responsibility for the research work as a whole
- The following cannot claim to authorship for a research publication without substantial contribution:
    o Department heads, other position of authority, or friends of authors;
    o Those who have provided routine technical contributions

- o Those who have provided routine assistance in some aspects of the research work

- o Funders

- o Supervisors of a research team

- All individuals (e.g., research students, research assistants, and technical writers) who have contributed to a research project but do not meet the criteria for authorship must be properly acknowledged.

10. Fees

DRC may charge data recipient for services to retrieve, process and provide the requested data, or any further costs for clerical and statistical support and delivery. The payment depends on the type of data access.

- For "Open Access data", no payment is required.
- For "Restricted Access Data", UoG staff shall be charged 10% of the total research budget while UoG students shall be charged 5% of the total research budget. External researchers (non-UoG employees) shall pay 15% of the total research budget to DRC.
- For "Closed Access Data", UoG staff are required to pay 15% of the total research budget while UoG students are required to pay 10% of the total research budget to DRC. External researchers (non-UoG employees) shall be charged 20% of the total research budget.
- For "Access that Needs Primary Data Collection", the researchers shall cover all the costs associated with data collection, management, and analysis. In addition, UoG staff are required to pay 15% of the total research budget while UoG students are required to pay 10% of the total research budget to DRC for administrative-related costs. External researchers (non-UoG employees) shall be charged 20% of the total research budget.

References

1. United Nations, *Handbook on Civil Registration and Vital Statistics Systems: Management, Operation and Maintenance*. 2018, New York, USA: UN.
2. Central Statistical Agency (CSA). *Quick Links: Central Statistical Agency*. 2020 28/11/2020]; Available from: https://www.statsethiopia.gov.et/.
3. Federal Ministry of Health (FMOH). *Ethiopian Health Analytics Platform*. 2020 28/11/2020]; Available from: http://www.moh.gov.et/ejcc/en/node/33.
4. Central Statistical Agency (CSA). *Specific Survey Titles*. 2020 28/11/2020]; Available from: https://www.statsethiopia.gov.et/our-survey-reports/.
5. National Institute of Health. *NIH Data Sharing Policy and Implementation Guidance*. 2020 03/11/2020   19/11/2020];   Available   from: https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
6. Research Data Alliance, *Data Sharing Principles in Developing Countries*, in *RDA at Open Data SSDC International Workshop*. 2014, Research Data Alliance: Nairobi, Kenya.
7. The University of Chicago. *Data-sharing Agreements*. 2020 20/11/2020]; Available from: https://ura.uchicago.edu/page/data-sharing-agreements.
8. Emerson, C., et al., *World Data System (WDS) Data Sharing Principles*. 2015, Zenodo: Geneva, Switzerland. p. 1.
9. Dixon, B.E., et al., *A vision for the systematic monitoring and improvement of the quality of electronic health data.* Stud Health Technol Inform, 2013. **192**: p. 884-8.
10. Dixon, B.E., et al., *Electronic health information quality challenges and interventions to improve public health surveillance data and practice.* Public Health Rep, 2013. **128**(6): p. 546-53.
11. Dixon, B.E., et al., *Extending an open-source tool to measure data quality: case report on Observational Health Data Science and Informatics (OHDSI).* BMJ Health Care Inform, 2020. **27**(1).
12. University of Pittsburgh. *Data Use Agreements*. 2020 [cited 2020 17/05/2020]; Available from: https://www.osp.pitt.edu/ccc-data-use-agreements.

Data Sharing and Use Agreement Template

Purpose of Agreement

The purpose of the data sharing agreement between the Dabat Research Center (DRC) and [Requesting Organization] is to provide Recipient with access to Dabat Research Center data for use in the following titled research project: [Project Name] under the direct supervision of [Principal Investigator] in accord with the Dabat Research Center Data Sharing Policy. [Describe the objectives and benefits the Requesting Organization hopes to achieve.]

Duration of Agreement

This agreement will commence at midnight on [DD/MM/YYYY]. This agreement will remain in place for [days/months/years/indefinitely] and will end on [DD/MM/YYYY] (if applicable) or until terminated by either party.

Description of Data

The Dabat Research Center is willing to provide health and demographic surveillance data in a format that will ensure anonymity and data safety to [Requesting Organization] where possible and as required. [Describe specific data being provided in this agreement. Include variables names, descriptions, format, level of security/sensitivity.]

Data Access

Data will be transferred from Dabat Research Center to the [Requesting Organization] by [email attachments/portable storage medium/data archive facility]. Information will be shared on a strict need-to-know basis only and the data will only be processed by individuals or groups of individuals in order for them to perform their duties in accordance with one or more of the defined purposes. [Include individuals or groups of individuals that will have access to the data.]

Terms of Agreement

1. The Dabat Research Center will provide the requested data to [Requesting Organization] for the defined purposes

2. The [Requesting Organization] will not release the individual names, addresses, or information that could be linked to an individual, nor will the recipient present the results of data analysis (including maps) in any manner that would reveal the identity of individuals.

3. The [Requesting Organization] shall ensure that the shared data is controlled through secure, authenticated mechanisms to prevent any unauthorized third-party access.

4. The [Requesting Organization] will not release data to a third party without prior approval from the data provider.

5. Any third party granted access to data, as permitted under condition #4, above, shall be subject to the terms and conditions of this agreement. Acceptance of these terms must be provided in writing by the third party before data will be released.

6. Data transferred pursuant to the terms of this Agreement shall be utilized solely for the purposes set forth in the "Purpose of Agreement" section.

7. The [Requesting Organization] will not share, publish, or otherwise release any findings or conclusions derived from analysis of data obtained from the data provider without prior approval from the data provider.

8. All data transferred to the [Requesting Organization] by the Dabat Research Center shall remain the property of the Dabat Research Center and shall be returned to the Dabat Research Center upon termination of the Agreements.

9. Neither the Dabat Research Center nor its collaborators 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 party.

10. The [Requesting Organization] must provide proper acknowledgement in all forms of publications produced using the Dabat Research Center shared data.

Termination of Agreement

This agreement may be terminated earlier, if found contrary to the data sharing policy or on the request of the data user to terminate, on the receipt of a month written notice, to be given by either party to the agreement.

Payment

The [Requesting Organization] shall pay a total of ETB _____ (___% of its project) to the Dabat Research Center for the data to be shared. The payment shall be completed not greater than a week following the execution of this agreement. The requested data will be released within two weeks after the [Requesting Organization] provide evidence for the payment.

Signatures

IN WITNESS WHEREOF, both the Dabat Research Center, through its duly authorized representative, and the [Requesting Organization], through its duly authorized representative, have hereunto executed this Data Sharing and Use Agreement as of the last date below written.

Dabat Research Center   **[Requesting Organization Name]**

_____   _____
Signature   Signature

_____   _____
Printed Name   Printed Name

_____   _____
Date   Date