Johns Hopkins Bloomberg School of Public Health

Institute for International Programs

COMSA

Countrywide Mortality Surveillance for Action

Mozambique

Project deliverable due February 1, 2017

O Data Access Module: Completed Data Access Strategy

Submitted to the

Bill & Melinda Gates Foundation

Grant # OPP1163221

March 15, 2017

COMSA Data Access Plan

***Note**: This document was adapted from the CHAMPS data access plan. Several sections have been derived from this plan.*

Document Revision History

Date	Reason For Changes	Version
2/1/2017	Initial version submitted to BMGF	1.0 Draft
3/15/2017	Post BGMF Review	2.0 Draft
5/26/2017	Updated by JHU with input from Mozambique team	2.1 Draft

Overview

Countrywide Mortality Surveillance for Action (COMSA) – Mozambique

COMSA is a three-year project aimed at improving the measurement and monitoring of mortality and cause of death in Mozambique. Better understanding of causes of death will build on Mozambique's previous success in reducing child mortality and help accelerate further mortality reduction to meet the Sustainable Development Goals.

COMSA will support Mozambique to develop and implement a sample registration system (SRS) of pregnancies, births and deaths, with cause of death assessment in the total population, with particular emphasis on children under-five using verbal autopsy and results from an innovative approach that relies on minimally invasive tissue sampling (MITS) implemented through a separate project on Child Health and Mortality Prevention Surveillance (CHAMPS) lead by the Emory University.

COMSA aims to achieve four primary outcomes:

(1) a sustainable and country owned sample registration system (SRS) for mortality and cause of death surveillance, enabling national and subnational comparison leading to action;
(2) improvements in measurement of cause of death among children under-five through use of MITS results implemented by the CHAMPS project to validate and improve verbal autopsy-based child cause of death;
(3) annual national and subnational mortality and cause-specific rates generated using the empirical SRS data but also through sound statistical modelling, combining data from the SRS, the CHAMPS' MITS results and other available national surveys such as Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Survey (MICS), leading to production of robust mortality and cause of death levels at national and subnational levels;

(4) a sustainable system for mortality and cause of death data collection, analysis and use in place in Mozambique, owned and run by the government, generating frequent national and subnational.

To achieve these results, COMSA will establish random sample of communities in each province of Mozambique and employ resident community-based worker, appointed by the communities themselves, to identify and report data on pregnancies, birth outcomes and deaths, as they occur in these communities. All deaths identified will be followed up by trained teams of interviewers to collect verbal and social autopsy data. Furthermore, COMSA will establish under-five mortality surveillance in two central hospitals in two provinces where a link will be made with the CHAMPS project. The hospital mortality surveillance will allow CHAMPS to carry out MITS on a sample of under-five deaths at these two sites. COMSA will also follow-up all under-five deaths in this sample to carry out verbal and social autopsy.

COMSA, therefore, will generate a wealth of data on mortality and cause of deaths that will be analyzed and released for public access. This Data Access Plan establishes rules and steps for data publication and access. The COMSA data will comprise:

(1) Pregnancies
(2) Births outcomes including miscarriages, stillbirths and live births
(3) Verbal autopsy on all deaths identified in the SRS and on a subsample from hospitals
(4) Data on MITS results produced by CHAMPS

Approach to Data Access and Sharing

Our goals for Data Access and Sharing are to encourage and facilitate wider use of the data generated by COMSA for improved understanding of national levels, trends and patterns of mortality and distribution of cause of death. By facilitating access to these data, COMSA will also promote an increased demand for these data both country level as well as at global level. To reach these goals, data must be promptly and broadly disseminated to meet the goal of understanding and reducing mortality. With support from the Bill & Melinda Gates Foundation we embrace the need to share high quality data in a timely manner so that we and our Mozambique partners such as the Ministry of Health (MISAU), the National Institute of Statistics (INE), the National Institute of Health (INS) and other stakeholders can better understand the trends and causes of mortality while developing more efficient and effective strategies, policies, and interventions to meet this challenge.

We will adopt a continuous improvement approach to timely information sharing by regularly exploring how we can make COMSA data more accessible and usable. We will ensure that COMSA data is accompanied with clear documentation intended to facilitate data use and enhance interpretability. Moreover, we will monitor and evaluate the best practices of other institutional data sharing efforts amongst our partners and peers around the world.

More importantly, we will strive to build local governance and management of COMSA data for the purpose of assisting country leaders and key partners in decision making.

Purpose of the Data Access Plan

This document defines the COMSA Data Access approach to promoting the transparent collection, processing, analysis and sharing of COMSA surveillance data, while ensuring privacy, availability, security

and governance. JHU will implement, in collaboration with Mozambique in-country partners, a comprehensive data access framework of practice that:

- Outlines data governance guiding principles;
- Details data governance structure;
- Defines the type and quality of data and information to be disseminated;
- Defines the contexts, conditions and criteria for accessing data; and
- Outlines the process for requesting, monitoring, and safeguarding data.

While COMSA focuses on providing transparent management, rapid dissemination, and clearly stated rules for access and use of COMSA data, national and international laws governing data use and sharing must be adhered to and enforced. As a result, this data access plan also addresses our approach to understanding the circumstances under which adherence to regulations on data use and management will apply for COMSA in reference to applicable laws, regulations and policies. Finally, this document serves as a guide to data governance, recognizing that strategies and standards will continue to evolve over time.

Data Governance Guiding Principles

The COMSA data governance guiding principles reflect those of Johns Hopkins University and the Bill & Melinda Gates Foundation:

1. In-Country Governance: Policies and procedures for data governance will strive to build in-country governance and reflect the values, beliefs and traditions of Mozambique. Governance policies should facilitate data management, access, and use by entities within Mozambique.

2. Confidentiality: Respect must be given to matters of identity, privacy, and confidentiality as they pertain to the individuals and communities from or about whom data are collected.

3. Attribution: Respect must be given to matters of attribution and recognition as they pertain to researchers, evaluators, and their collaborators.

4. Accountability: All processes and procedures for data access will be transparent, clear, and consistent with data management standards that ensure quality data, appropriate security and equitable access.

5. Stewardship: All who produce, share and use data are stewards of that data. They share responsibility for ensuring that data are collected, accessed and used in appropriate ways, consistent with applicable laws, regulations and current and evolving international standards of ethical research conduct.

6. Innovation: Data access encourages diversity of analysis and opinion while assuring, through detailed documentation, consistent understanding of the nature of data collection and scope and limitations of data; facilitates the evaluation of alternative hypotheses; permits meta-analyses.

7. Efficiency: Providing widespread access to datasets enabling application of energies and limited resources towards the most productive outcomes stimulating new research endeavors.

8. Collaboration: Ensuring access to data among institutions and across disciplines will enable greater productivity and creativity. COMSA will make every effort to collaborate with key stakeholders in data access governance policies decisions.

In addition, consistent with the Bill & Melinda Gates Foundation's Open Access Policy, underlying data sets associated with publications in peer-reviewed journals, will be made immediately and freely available, subject to removal of any protected health information or other personally identifiable information and in accordance to local laws and regulations.

Data Governance Structure and Approach

COMSA data and analytical outputs will be governed jointly by JHU and INE and INS. These three institutions will put in place a Data Oversight Team (DOT). The DOT will issue guidance, establish standards, prioritize initiatives, and develop policies that ensure consistent data management, curation, and dissemination. Furthermore, the DOT will be expected to review data usage monitoring reports/indicators, approve change requests, deliberate risk mitigation plans, and issue guidance regarding data sharing issues or challenges.

The team will consider specific policy issues including, but not limited to:

- adhering to the housing of data within specific in-country institution or cloud geo-location;
- data privacy and identifier concerns, data sharing / reuse / publication issues;
- processes and data structures to support identity linkage and longitudinal data management, analysis and publication clearance;
- criteria and policies for evaluation and approval of data access requests; and
- concerns for data timeliness, quality, and completeness.
- collaborating with the CHAMPS team to maintain interoperability and comparability standards

COMSA will continue to evolve the data governance approach as needed to optimize the process and to achieve and sustain the goal of open data access.

Levels/Types of COMSA Data

Data produced by COMSA will require the collection, transfer, storage, and dissemination of sensitive personal and health information. Utilizing data privacy and confidentiality guidelines, a set of four (4) distinct levels of COMSA data have been established in order to facilitate the governance, confidentiality, privacy, security, and ethical exchange of COMSA data (Table 1).

Table 1: Definition of COMSA Data Privacy Levels/Types

Level	Data Type	Description
Level 1	**Summarized Dataset**	Inclusive of high-level summaries of operational or aggregated case activity, that excludes any identifiable information. Data Categories by year, age-group (<28 days, 28days-59mo, 5-11 years, 12+ years), and province: ● Pregnancy outcomes ● Deaths ● Cause-specific mortality fraction and rates ● Number of VASA's completed ● Number of MITS consented
Level 2	**Case-level De-Identified Dataset** [1]	Identifiers of individuals (or their respective relatives, employers, or household members) are removed such that the information could not be used alone or in combination with other information to identify an individual who is the subject of the information. Age related dates will be converted to age-group (<28 days, 28days-59mo, 5-11 years, 12+ years). By definition, a "de-identified" dataset cannot be used in combination with other data to identify individuals. Thus, a dataset that may be leveraged in combination with another source makes it a Level 3 – Limited Dataset.
Level 3	**Case-level Limited Dataset** [2]	A limited set of identifiable patient/subject information including actual dates but excluding other potentially identifying data as described in Appendix 2. A "*limited data set*" of information may be disclosed to an external party without a subject's authorization if certain conditions are met: ● First, the purpose of the disclosure may only be for research, public health or health care operations. ● Second, the person/entity receiving the information must sign a Data Use Agreement.[3]
Level 4	**Identified Data**	Dataset is inclusive of any potentially identifying patient/subject information provided to the Program Office (subject to informed consent constraints). With the exception of names, potentially personally identifying information may be disclosed to an external party without a subject's authorization if certain conditions are met: ● First, the purpose of the disclosure may only be for research, public health or health care operations. ● Second, the person/entity receiving the information must sign a Data Use Agreement.[3] ● Specific examples of identifiable information include geo-location coding, national identifiers, or DSS identifiers. Additional details provided in appendices.

[1] See Appendix 1 for detailed description of De-Identified Dataset

[2] See Appendix 2 for detailed description of Limited Dataset

[3] Data use agreements must be signed with the appropriate governing institutions. Initially, Johns Hopkins and INE will share this responsibility. This may change as the study progresses.

Context of Data Access and Exchange

Programmatic Operational Data Access and Exchanges

Programmatic operational data access and exchanges are the continual exchange of surveillance and operational data between JHU and COMSA in-country collaborators (INE and INS) for the purposes of source data curation, analysis, quality assurance, and continuous programmatic evaluation. Operational data exchanges inherently involve all four levels or types of COMSA data.

Open Data Sharing and Dissemination of Curated Surveillance Data

Open data sharing and dissemination is the provisioning of high quality COMSA surveillance data, either by open publication or by direct fulfillment of specific requests, to persons/entities that are not essential to the programmatic operations of COMSA.

Data Access and Exchange Governance Matrix

It is the intention of COMSA to facilitate open and timely access to all levels (or types) of COMSA surveillance data without compromising the integrity of our guiding principles or the privacy of the individuals who have consented to surveillance activities. Therefore, fulfillment of open data requirements/requests necessitates adherence to specific access methods, procedures, governance standards, and agreements related to the access and use of data. The terms, conditions, timing, and content of data openly shared or disseminated by COMSA to persons/entities are governed by the policies and procedures established by the COMSA DOT.

Table 2: Data Access and Exchange Governance Matrix

Level	Data Type	Access Method	Governance	Required Agreement(s)	Frequency of Data Refresh	Restrictions, Limitatio
Level 1	Summarized Dataset	Publically available for immediate consumption (via website/portal)	Monitoring of Consumption (downloads, page-views, citations)	End User License Agreement	Updated every 120 days (120 days refers to the period from date of death to posting to public site)	Unrestricted Open to the General Limitation(s): To ensu obfuscation methods may be employed to timing
Level 2	Case-level De-Identified Dataset	Registration (via website/portal)	Monitoring of Consumption (registration metrics, downloads/transmissions, citations) Pre-publication Review (when applicable)	End User License Agreement	Updated every 120 days (Deaths will require up to 120 days for complete data collection and posting to public site)	Minimally Restricted Access to Case-level D limited to users who terms and conditions Limitation(s): Depend statistical obfuscation methods may be emp or event timing to en
Level 3	Case-level Limited Dataset	Application (via website/portal)	Application Review & Approval	Data Use Agreement (DUA) Accredited IRB Approved Protocol	Defined by application request and terms of Data Use Agreement	Moderately Restricte Limitation(s): Depend statistical obfuscation methods may be emp or event timing to en
Level 4	Identified Data	Application (via website/portal)	Application Review & Approval Accredited IRB Oversight	Data Use Agreement (DUA) Accredited IRB Approved Protocol	Defined by application request, specifics of the IRB approved protocol, and the terms of the Data Use Agreement	Highly-Restricted

Data Sharing Requests & Monitoring

The DOT is responsible for the establishment of data sharing policy and the monitoring of sharing reports and indicators.

The turnaround times for fulfillment of data access requests and approvals are dependent on the level (or type) of dataset requested, the complexity of the data requested, and the method of dissemination. Standardized Level 1- 2 datasets should be immediately accessible. Custom requests of Level 1-2 data (i.e., non-standard, complex, or large image sets) may require up to six-weeks for fulfilment. Level 3-4 applications should be evaluated within six weeks, and fulfilment times may vary depending on the complexity of the order and/or the terms of the data use agreements

Publically Available Level-1 Datasets

Access to COMSA Level 1 datasets will be unrestricted and open to the public. Users will not be required to register with COMSA. However, they will be required to agree to the terms of an End User License Agreement prior to downloading any content. COMSA will request the appropriate citation of any open data or materials utilized for presentations or publication. Datasets and materials should be made immediately for consumption via the COMSA website.

COMSA will utilize common web monitoring tools (e.g., Google Analytics) to track industry standard metrics regarding page-views and frequency of static content/standard dataset downloads. Reports will be made accessible/distributed to the appropriate Working Groups and the DOT per standard operating procedures.

Registered Access to Level-2 Datasets

To access Level 2 datasets, users will be required to register with COMSA and agree to the terms of the End User License Agreement. In addition to requesting the appropriate citation of any open data or materials utilized for presentations or publication, COMSA will request the opportunity for the prepublication review of any publications derived from Level 2 COMSA data. Standardized datasets or self-service reports should be immediately available for consumption via the COMSA website.

The DOT will review reports describing the frequency and type of registrations and the datasets downloaded per standard operating procedures. Data requests, downloads of registered users will be logged

Whenever feasible and applicable, an accounting of scholarly citations generated by registered users utilizing COMSA data will be monitored.

Application for Level-3 and Level-4 Datasets

Access to Level 3 and Level 4 datasets will require the submission of a detailed application with COMSA. Applications will be reviewed and either approved or denied by the DOT. Upon approval, the applicant is required to execute a Data Use Agreement (DUA). Requests for Level-4 datasets may be subject to IRB Oversight and therefore require an IRB approved protocol.

Data Dissemination

Initial Phase-in and Availability Thereafter

The COMSA approach to quickly disseminating data may be challenging when balanced with a need for accuracy and reliability in the newly accumulated surveillance information.

In an effort to reach a reasonable accord, we are proposing an initial "phase-in" period to allow for preliminary analysis and interpretation of the data. The phase-in period is effective for up to 6 months from the completion of the first date of data collection in a phase I provinces and applies to the group of phase I provinces. The phase-in period is not meant to apply to each individual province. During the phase-in period, access to the curated (Level 2-4) surveillance data from each case is only available to JHU, INE and INS. Summary level information (Level 1) is reportable within the context of case counts or other aggregate-level metrics.

After the phase-in period has elapsed, all levels of curated case-data from the initially deferred cases would be openly available. Subsequent curated case data would then be made available, in their entirety, within 120 days of data collection.

At any time, including the phase-in period, data and information supporting the reporting of notifiable conditions will be made immediately available through appropriate reporting mechanisms to the MISAU and INS. Release of information to support case reporting requirements for a specific case could precede final autopsy determinations to allow for public health action prior to release of fully curated COMSA case data.

As methodological approaches and technique changes occur over time, it is anticipated that COMSA, after consideration of potential impact of the advances/changes, will reopen applicable prior cases and perform additional analysis and/or generate refined cause of death determinations. The timeline for release of these new findings will depend on the operational and analytic approaches required to generate findings for individual cases and aggregate data sets. All data dissemination from COMSA is subject to constraints defined within Material and Data Transfer Agreements, Protocols and Informed Consents.

Dissemination Methods

The designated data stewards within COMSA are responsible to the fulfillment of sanctioned data requests. A number of data dissemination methods are possible to facilitate open and timely data sharing of COMSA Surveillance data:

- Reports, Publications, and/or Visualizations
- Standardized Datasets (or Packages)
- Custom Data Request Fulfillment

The data type may influence or restrict the methods of dissemination, and are therefore subject to the policies and procedures sanctioned by the COMSA DOT (See Data Access Exchange and Governance Matrix). Data dictionaries and data standard references (if applicable) will be accessible/provided whenever applicable.

Data Formats

COMSA data will be made available one or more of the following formats

- CSV
- PDF (e.g., graphics, image, maps)
- XML

Storage and Retention

Data Storage

- Data will be stored in the COMSA database, hosted on a cloud-based server. The data will be accessible via password- and role-based controls.
- "Grass-roots" data collection will occur via electronic data capture at the source on tablets or phones or in rare cases on paper with subsequent abstraction to an electronic data capture system. As a result, unprocessed COMSA data will temporarily exist on electronic devices or paper. The data collection Supervisor responsible for data entry coordination and validation must ensure the transfer of paper type data to COMSA server within a specified time frame. It is anticipated this will be within 2 weeks from when the paper tools are collected.

Data Retention

It is the intention of the COMSA to retain surveillance data for at least the duration of the project.
COMSA data retention policies and procedures are subject to the specific and distinct contractual obligations between JHU and in-country Collaborators. However, retention of datasets shared or disseminated by JHU are subject to the specific terms of the agreements established at the time of data access fulfillment.

Data Security & Privacy

As good stewards of data and respectful of local, international and US regulations, COMSA will model its data management, security and access practices with guidance provided by Johns Hopkins University IRB policy and laws and practices of Mozambique and the United States. JHU and in-country Collaborators will apply stringent physical security controls and processes to safeguard data and protect the privacy of COMSA subjects.

JHU and in-country Collaborators will continuously work to ensure open access to the global community as well as appropriate data management strategies. We will continuously emphasize our commitment to ensuring that COMSA surveillance data are promptly and broadly accessible, while sufficiently protected

to meet the ultimate goal of enhancing global understanding of the causes of mortality, driving programmatic and research priorities towards reducing mortality.

Future Storage and Retention

It is the intentions of both JHU and the Bill and Melinda Gates Foundation (BMGF) that open access to COMSA data persist in perpetuity for the benefit of future reference, research, analysis, and discovery so long as it is feasible for either party to facilitate access and/or management of dissemination.

This section intends to address the future retention, administration, and dissemination of COMSA data when programmatic administration of the COMSA is wholly absorbed by the Mozambique government or transferred to the BMGF (or their designee) due to either:

(a) Completion of the Terms of COMSA Program Terms and/or Objectives

(b) Transfer of COMSA Programmatic Administration from JHU

In either scenario, all curated data as well as the subsequent responsibilities of data stewardship of that data are to be transferred to the BMGF (or their designee) in accordance with the terms outlined in applicable termination agreements. Henceforth, COMSA data access and dissemination will be subject to terms and policies as defined by the BMGF (or their designee). Moreover, these terms may be subject to the discretion of the in-country partners or the ministries of health.

Previously released COMSA data (Levels 1 - 4) to COMSA partners will be subject to the respective terms of each data use agreement and/or end user license agreement.

The original COMSA grantee, JHU, shall retain data access rights to data generated during its term of program administration pursuant to the explicit terms outlined in applicable termination agreements.

Appendixes

The following appendices describing *De-identified* and *Limited Data* are representative of current US practice and policy standards. The intent of these representations are to establish a foundation for discussion and refinement amongst COMSA stakeholders and the COMSA DOT. The ultimate outcome envisioned are internationally adapted definitions that harmonize the perspectives and practices of our Sites across countries in Africa and South Asia.

Appendix 1 – Definition of De-Identified Data

The following identifiers of the individual or of relatives, employers or household members of the individual must be removed:

- Names;
- All geographic subdivisions smaller than a State/Province (*or comparable country-specific subdivisions*), including street address, city, county, precinct, and their equivalent geographic coordinates.
- All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 (*or alternative per country/region*) and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 (*or alternative per country/region*) or older; Age categories will be (<28 days, 28 days-59 mo, 5-11 years, 12+ years).
- Telephone numbers;
- Fax numbers;
- Electronic mail addresses;
- National/DSS ID;
- Medical record numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate/license numbers;
- Vehicle identifiers and serial numbers, including license plate numbers;
- Device identifiers and serial numbers;
- Web Universal Resource Locators (URLs) and Internet Protocol (IP) address numbers;
- Biometric identifiers, including finger and voice prints;
- Full face photographic images and any comparable images; and
- Any other unique identifying number, characteristic or code.

Appendix 2 – Definition of a Limited Data Set

While COMSA data will be collected on individuals outside the US, the COMSA team has aligned policies for research subject privacy with US regulations to ensure appropriate levels of protection for subject data. In addition to summary and de-identified case-level data, COMSA will also provide data in a "limited data set" defined as a limited set of identifiable patient/subject information. The use of limited data sets will enable analysis that may be difficult or impossible using fully de-identified data while providing minimal exposure of potentially identifying information.

All the following identifiers must be removed in order for health information to be a "limited data set":

- Names;
- Street addresses (other than town, city, state, zip code, (*or comparable country-specific subdivisions*);
- Telephone numbers;
- Fax numbers;
- E-mail addresses;
- National/DSS ID;
- Medical records numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate license numbers;
- Vehicle identifiers and serial numbers, including license plates;
- Device identifiers and serial numbers;
- URLs or IP address numbers;
- Biometric identifiers (including finger and voice prints);
- Full face photos (or comparable images); and
- Household specific geographic coordinates.

The health information that may remain in the information disclosed includes:

- Dates such as admission, discharge, service, DOB, DOD;
- City, county state, five digit or more zip code (*comparable country-specific subdivisions);* and
- Ages in years, months or days or hours.

It is important to note that this information is still protected health information. It is not de-identified information and is still subject to the requirements of applicable Privacy Regulations.