



HHS Public Access

Author manuscript

Ann Epidemiol. Author manuscript; available in PMC 2020 August 22.

Published in final edited form as:

Ann Epidemiol. 2020 June ; 46: 1–4. doi:10.1016/j.annepidem.2020.04.001.

Ensuring Ethical Data Access: The Sierra Leone Ebola Database (SLED) Model

Yelena Gorina, M.P.H., M.S.¹, John T. Redd, M.D., M.P.H.², Sara Hersey, M.P.H.³, Amara Jambai, M.D.⁴, Peter Meyer, M.P.H., M.A.⁵, Ansumana S. Kamara, M.S.⁶, Alimamy Kamara, M.B.A.¹, Jadhah D. Harding, B.S.⁷, Brima Bangura⁷, Mohamed A.M. Kamara⁷

¹Centers for Disease Control and Prevention, United States

²The Department of Health and Human Services, Office of the Assistant Secretary for Preparedness and Response, United States

³World Bank

⁴Sierra Leone Ministry of Health and Sanitation, Sierra Leone

⁵NORC at The University of Chicago

⁶African Field Epidemiology Network (AFENET), Sierra Leone

⁷Sierra Leone Ebola Database Data Team, ICAP, Sierra Leone

Abstract

Background—Organizations responding to the 2014-2016 Ebola epidemic in Sierra Leone collected information from multiple sources and kept it in separate databases, including distinct data systems for Ebola hot line calls, patient information collected by field surveillance officers, laboratory testing results, clinical information from Ebola treatment and isolation facilities, and burial team records.

Methods—Following the conclusion of the epidemic, the Sierra Leone Ministry of Health and Sanitation (MoHS) and United States Centers for Disease Control and Prevention (CDC) partnered

Corresponding Author: Yelena Gorina, Office of Analysis and Epidemiology, National Center for Health Statistics, Centers for Disease Control and Prevention. 3311 Toledo Road, Hyattsville, MD 20782, USA, yag9@cdc.gov, Telephone: 301-458-4241. Yelena Gorina: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing; John T. Redd: Conceptualization, Methodology, Writing – original draft; Writing – review & editing, Supervision; Sara Hersey: Conceptualization, Methodology, Writing – review & editing, Supervision; Amara Jambai: Conceptualization, Methodology, Writing – review & editing, Supervision; Peter Meyer: Conceptualization, Methodology, Writing – original draft, Supervision; Ansumana S. Kamara: Methodology, Investigation, Writing – review & editing; Alimamy Kamara: Methodology, Software, Writing – review & editing; Jadhah D. Harding: Investigation, Writing – review & editing; Brima Bangura: Investigation, Writing – review & editing; Mohamed A.M. Kamara: Investigation, Writing – review & editing

Publisher's Disclaimer: This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Publisher's Disclaimer: Disclaimer: The conclusions and opinions expressed by authors contributing to this journal do not necessarily reflect the official position of the U.S. Department of Health and Human Services, the Centers for Disease Control and Prevention, or the authors' affiliated institutions.

Competing interests

The authors have no competing interests to declare.

to collect these disparate records and consolidate them in the Sierra Leone Ebola Database (SLED).

Results—The SLED data will be used to provide a lasting resource for post-epidemic data analysis and epidemiologic research, including identifying best strategies in outbreak response, and to help families locate the graves of family members who died during the epidemic.

Conclusion—This report describes MoHS and CDC processes to safeguard Ebola records while making the data available for public health research.

Keywords

Ebola virus disease; ethics; privacy; data ownership; data sharing; data access; Sierra Leone Ebola Database; SLED

Background

The large-scale 2014–2016 Ebola virus disease (Ebola) epidemic in Sierra Leone demonstrated the importance of coordinated efforts in data collection and management¹. Data standardization is crucial for supporting and facilitating epidemic response. It is also vitally important for post-epidemic data analysis and epidemiologic research to help plan for future responses².

To prevent further spread of the disease, from October 2014 through November 7, 2015 all burials in Sierra Leone were mandated to be conducted by specially trained and equipped burial teams³. In 2015, the Sierra Leone Ministry of Health and Sanitation (MoHS), the owner of the data collected in Sierra Leone, in collaboration with the United States Centers for Disease Control and Prevention (CDC), set a goal of collecting, consolidating, and linking Ebola data with a primary objective to help families locate the grave of their loved ones.

The international non-governmental humanitarian organization Concern Worldwide supervised and recorded more than 16,000 safe and dignified burials in two cemeteries in the Western Area⁴. Using burial records, the organization's family liaisons helped 1,473 family members locate the graves of their loved ones (Otieno D., Concern Worldwide: personal communication). Unfortunately, for many families who were not able to be present at the funeral of their loved ones, grave locations are still not known.

During collaborative data collection efforts with epidemic response organizations, MoHS and CDC realized the greater potential of a larger, consolidated Ebola database, including epidemiological investigation and clinical records data in addition to burial records. By the end of 2016, 68,000 (100%) burial records, 97,588 (100%) of epidemiologic investigation records, 106,172 (100%) laboratory testing records, 239,858 (about 99%) Ebola hotline alert records, and 7,245 (about 80%) Ebola facility records were consolidated into a data collection that is now referred to as the Sierra Leone Ebola Database (SLED). Each of the SLED data categories has its own limitations: The alert call center was established in September and burial teams in October of 2014, three and four months, respectively, after the Sierra Leone epidemic was declared. Laboratory testing records were not received

subsequent to October 2015. Finally, some Ebola facility records were not available for collection. In addition, none of the SLED components include information about hidden³ or asymptomatic cases⁵. Despite these limitations, the SLED data became a unique consolidated data source for the Sierra Leone Ebola epidemic. Once the database was established, MoHS and CDC recognized the need to develop an ethically appropriate mechanism for researchers' secure data access that will protect personally identifiable information (PII) while preserving MoHS's data ownership permanently.

The solution came from an analysis of the best international practices of data ownership and privacy protection⁶⁻¹⁷ and discussions with staff from the U.S. National Center for Health Statistics' Research Data Center¹⁸ (RDC) who have extensive experience in enabling access to potentially identifiable data in a secure environment¹⁹.

As Luciano Floridi wrote in 1999, "privacy, accuracy, intellectual property and access, ... also security and reliability, ... have been so transformed by the computing technology in which they are embedded that they acquire an altered form and new meanings.. we need a conceptual interface to apply ethical theories to new scenarios."²⁰ A significant body of work exists on developing a unified big data ethics framework²¹⁻²⁵. In developing SLED, CDC and MoHS made efforts to assure that the data, collected by multiple organizations, could be consolidated and made available to researchers. The process is consistent with major principles of preserving data confidentiality and MoHS data ownership and custody of the data. On behalf of the MoHS, the RDC provides access to high utility data and prevents unauthorized use or disclosure of the information.

The SLED Model

The core of the developed solution includes: supporting, training, and mentoring an incountry team of Sierra Leonean data managers (SLED Data Team); developing a research proposal submission and approval process; facilitating protected access to the SLED data for researchers via the RDC; and supporting health research in Sierra Leone. The MoHS' ownership of the SLED data is assured by hosting and processing the data in the secure MoHS location by the SLED Data Team.

Researchers access the SLED data through the National Center for Health Statistics RDC, which was established in 1998 to provide a mechanism whereby researchers could access data files in a secure environment without jeopardizing the confidentiality of respondents and providing protection of the integrity of the data themselves. This is done by limiting researchers' access location to the physical data laboratories where RDC staff can prevent the removal of data. The physical laboratories contain servers isolated from devices that can be used to transmit data. Users with approved by MoHS proposals can access the required data set but cannot take the data out of the environment. Additionally, a remote execution system works by accepting code from researchers, running the code against an approved data set, and then returning the output to the requester. The remote execution system allows for data analysis to be conducted without the researcher having to be at the physical RDC location.

Researchers can submit the programming code and receive the results over a secure RDC File Transfer Protocol (FTP) site (Figure 1). No data can be downloaded onto the researcher's computer or other device. The RDC strives to maintain confidentiality and prevent disclosure, but researchers with approved projects also have important responsibilities in preventing disclosure^{18, 19}. It is a researcher's responsibility to receive institutional approval to uphold ethical principles and guidelines. The RDC is responsible for preventing the loss of the data and advising the SLED Data Team on best practices for disclosure limitation. The RDC does not comment on scientific merit or impose any merit-based publishing guidelines,¹⁹ and researchers will be able freely and independently to publish results derived from SLED. Although researchers will not be able to share the data accessed via RDC, publishing and sharing data dictionaries, algorithms, definitions, merging procedures, and results will allow reproducibility of the results by other researchers.

Maintaining SLED data security and confidentiality

The SLED Data Team maintains as a guiding principle the security and confidentiality of the data it safeguards, especially PII. Security may be defined as preventing unauthorized release of PII and loss or unauthorized access, destruction, use, modification, or disclosure of the data^{26, 27}. Measures to ensure the security of data include providing both physical security (e.g. a guarded, physically secure MoHS location for data storage with an uninterrupted power supply) and electronic security (e.g. strong passwords, data encryption, and data backup). The SLED Data Team maintains data confidentiality by ensuring that PII is not released without the consent of the patient or a family member of the deceased. To reinforce these requirements, the SLED Data Team developed standard operating procedures to protect the confidentiality of all SLED case reports and files, and the SLED Team members undergo regular training on data confidentiality and security.

The research proposal approval process

A prospective investigator submits a proposal to the MoHS and SLED Data Team via the SLED-dedicated page on the NCHS RDC web-site²⁸ containing the list of the Ebola responding organizations that contributed data to SLED and data dictionaries for the SLED files. The proposal must contain the following information:

- a. Investigators' names, affiliations, positions, nationality (Sierra Leonean or not), and curricula vitae;
- b. Research questions, methods, references, and expected outcome;
- c. Geographic area and dates of the requested data;
- d. Requested SLED files and desired variables;
- e. Description of the non-SLED files and variables, if linkage is requested.

The NCHS RDC analyst sends proposals to the SLED Data Team who review the proposal for data availability, data confidentiality, proposal completeness, and public health relevance. Once complete, the proposal is forwarded to MoHS for approval. Revisions will be requested if needed. If the MoHS approves the proposal, the NCHS RDC Analyst notifies

the researchers that they must complete a confidentiality orientation, pass a post-orientation test with a score of 100%, and sign a written confidentiality agreement (Figure 2).

Currently, CDC funds RDC services for MoHS-approved projects submitted by any Sierra Leonean researchers and researchers from Ebola responding organizations that contributed data to SLED. Other research projects are subject to the regular RDC fee.

Preparing the data package

An RDC analyst completes a final screening of the approved proposal for data confidentiality and transfers it on to the SLED Data Team in Freetown, Sierra Leone. For each approved research project, the SLED Data Team extracts or derives data elements requested in the proposal, limiting them to the geographic and temporal scope of the project. Indirect PII (e.g. date of hospitalization) are replaced by derived variables (e.g. month/year) to decrease the possibility of disclosure.

After the data package is reviewed by the SLED Data Team supervisor, it is transferred to the RDC via a secure FTP site. The RDC analysts communicate with the researchers to facilitate data processing and assure data security (Figure 1).

Supporting health research in Sierra Leone

An important principle of SLED is to maintain access by African researchers, particularly from Sierra Leone. Preliminary testing showed that the RDC is sufficiently accessible from Sierra Leone and other countries in the region via the Internet. To accommodate SLED needs, the RDC expanded its range of programming software to include SPSS software, version 23.0 (IBM). The standard programs that are provided in the RDC are SAS software, version 9.4 (SAS Institute), Stata/MP-64 software, version 14 (StataCorp), and R software, version 3.3.3 (Lucent Technologies). The SLED Data Team and RDC analysts are available to assist Sierra Leonean researchers with the process of proposal application and setting up programming. To familiarize Sierra Leonean public health researchers with the secure data access concept, SLED data, RDC, and principles of data confidentiality and security, MoHS and CDC conducted a 3-day workshop in January 2018 attended by more than 90 participants from Sierra Leonean governmental and academic institutions and 2-day session in May 2019 attended by 171 participants.

Conclusion

In October 2017, the SLED secure data access concept was approved by the MoHS. The first phase of the SLED secure data access was released in Fall 2018. The Sierra Leone MoHS maintains its ownership and custody of the SLED data. The SLED page on the NCHS RDC website will provide user support with a description of the data, an application package, and guidance on the approval process.

The SLED Data Team and RDC are responsible for protecting the confidentiality of the patients or institutions while providing access to the SLED data for statistical purposes. Big data ethical principles of SLED data access include protection from statistical and country-

sensitive cultural information disclosure. These principles are maintained by employing both technical and behavioral techniques to protect sensitive information.

The technical techniques include producing minimal data extracts to answer specific research questions and the use of data access from the physical data laboratories or a remote execution analytic system. The extracts provide high research utility without releasing more data elements than are needed. The remote execution is a secure data access technology that allows an approved researcher to write code that is sent to the RDC and run on the extract. When the operation is completed, the output is returned to the researcher. This aggregate form of data allows the researcher to conduct analyses without ever seeing or obtaining the underlying microdata. The MoHS maintains data ownership and the SLED data do not leave MoHS premises.

The behavioral or social techniques include educating users about the importance of confidentiality, the review of proposals to access the data, and the periodic review of aggregate data in the output. The RDC requires that approved users take confidentiality training with a post-test and make sure researchers and his/her institution are aware of their responsibility to maintain ethical principles of the research. The SLED Team maintains the best possible quality of the data by preparing, verifying, and documenting data package.

In 2015, an international workshop convened in South Africa focused on the benefits of and barriers to sharing research data in order to improve public health²⁹. The participants emphasized the need for maintaining data privacy and confidentiality, the role of the data-contributing nation in data governance, and the importance of strengthening the role of national researchers via capacity building, collaboration, and receiving recognition. CDC and MoHS designed the SLED secure data access concept following these principles. Despite multiple challenges in developing and setting up SLED data access model, we hope that it may serve as a prototype for future secure and ethical data sharing.

Acknowledgments

The authors wish to acknowledge multiple organizations that participated in the Ebola Response in Sierra Leone and contributed the records to the Sierra Leone Ministry of Health and Sanitation and Sierra Leone Ebola Database. Special thanks go to the Sierra Leone Ministry of Health and Sanitation, the District Health Management and Ebola Response Teams, and the men and women of Sierra Leone whose work as surveillance officers, care providers, burial team members, alert phone operators, data managers, social mobilizers, and community leaders and organizers was the major strength that stopped the epidemic. We want to acknowledge The Catholic Agency for Overseas Development, Catholic Relief Services, Concern Worldwide, the International Federation of Red Cross and Red Crescent Societies, the International Rescue Committee, and World Vision for contribution of burial team records; e-Health Africa and the District Alert Call Centers for contribution of alert records; Doctors Without Borders (MSF Holland), GOAL, the International Federation of Red Cross and Red Crescent Societies, International Medical Corps, The 34 Military Hospital (Freetown), Partners in Health, and Save the Children for contribution of medical records; the Sierra Leone Central Public Health Reference Laboratory, the Chinese Center for Disease Control and Prevention laboratory, the Dutch Laboratory in Princess Christian Maternity Hospital, the Italian National Institute of Infectious Diseases/Emergency laboratory in Goderich, the MoHS Ebola laboratories, Nigerian laboratories, Public Health England laboratories, and Public Health Agency of Canada laboratories for laboratory testing records; and the US Centers for Disease Control and Prevention for contribution of the laboratory testing and epidemiological records and overall support of the project.

References

1. McNamara LA, Schafer IJ, Nolen LD, Gorina Y, Redd JT, Lot T, et al. Ebola surveillance — Guinea, Liberia, and Sierra Leone. *MMWR Suppl* 2016;65(Suppl-3):35–43. [PubMed: 27389614]
2. Marston BJ, Dokubo EK, van Steelandt A, Martel L, Williams D, Hersey S, et al. Ebola response impact on public health programs, West Africa, 2014–2017. *Emerg Infect Dis.* 2017 Suppl. (10.3201/eid2313.170727.)
3. Nielsen CF, Kidd S, Sillah AR, Davis E, Mermin J, Kilmarx PH. Improving burial practices and cemetery management during an Ebola virus disease epidemic - Sierra Leone, 2014. *MMWR Morb Mortal Wkly Rep.* 2015 1 16;64(1):20–7. [PubMed: 25590682]
4. Concern Worldwide. Annual report and financial statements 2015 (<https://www.concern.net/resources/annual-report-and-financial-statements-2015>.)
5. Mafopa NG, Russo G, Wadoum REG, et al. Seroprevalence of Ebola virus infection in Bombali District, Sierra Leone. *J Public Health Afr.* 2017 12 31;8(2):732. [PubMed: 29456826]
6. Brack M, Castillo T. Data Sharing for Public Health: Key Lessons from Other Sectors. Chatham House Centre on Global Health Security Research Paper; London: The Royal Institute of International Affairs 2015 (http://www.chathamhouse.org/sites/files/chathamhouse/field/field_document/20150417DataSharingPublicHealthLessonsBrackCastillo.pdf.)
7. Dyke SO, Dove ES, Knoppers BM. Sharing health-related data: a privacy test? *NPJ Genom Med.* 2016 8 17;1(1):160241–160246. [PubMed: 27990299]
8. Evans BJ. Much ado about data ownership. *Harvard Journal of Law & Technology* 2011, Volume 25, Number 1 pp. 70–130.
9. Federal Committee on Statistical Methodology. Statistical Policy Working Paper 22 (second version 2005). Report on statistical disclosure limitation methodology, statistical and science policy. Office of Information and Regulatory Affairs, Office of Management and Budget 2005 (<https://www.hhs.gov/sites/default/files/spwp22.pdf>.)
10. Kostkova P, Brewer H, de Lusignan S., et al. Who Owns the Data? Open Data for Healthcare. *Front Public Health.* 2016; 4: 7 [PubMed: 26925395]
11. Modjarrad K, Moorthy VS, Millett P, et al. Developing global norms for sharing data and results during public health emergencies. *PLoS Med.* 2016 1;13(1).
12. Moorthy VS, Roth C, Olliaro P, et al. Best practices for sharing information through data platforms: establishing the principles. *Bull World Health Organ.* 2016 4 1; 94(4): 234–234A. [PubMed: 27034512]
13. Moon S, Sridhar D, Pate MA, et al. Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola. *The Lancet.* 2015;386:2204–21
14. Oderkirk J, Ronchi E, Klazinga N. International comparisons of health system performance among OECD countries: opportunities and data privacy protection challenges. *Health Policy.* 2013 9;112(1–2):9–18. [PubMed: 23870099]
15. van Panhuis WG, Paul P, Emerson C, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health.* 2014 11 5; 14:1144. [PubMed: 25377061]
16. Stansfield S Who owns the information? Who has the power? *Bull World Health Organ.* 2008 3; 86(3): 170–171. [PubMed: 18368202]
17. Bygrave LA. Privacy and Data Protection in an International Perspective. *Scandinavian Studies in Law,* 2010: 165–200.
18. Centers for Disease Control and Prevention. National Center for Health Statistics. Research Data Center website (<https://www.cdc.gov/rdc/index.htm>.)
19. Centers for Disease Control and Prevention. National Center for Health Statistics. Research Data Center Publication Guidelines (<https://www.cdc.gov/rdc/b6pubeyond/pub600.htm>).
20. Floridi L Information ethics: on the philosophical foundation of computer ethics. *Ethics and Information Technology,* 1(1): 37–56, 1999.
21. Mittelstadt BD, Floridi L. The Ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics.* 2016 4;22(2):303–41. 2016. [PubMed: 26002496]

22. European Economic and Social Committee. The ethics of big data. 2017 Available from <https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data>.
23. Hand DJ. Aspects of data ethics in a changing world: where are we now? *Big Data*. 2018 9 1;6(3):176–190. 2018. [PubMed: 30283727]
24. Zook M, Barocas S, Boyd D, et al. Ten simple rules for responsible big data research. *PLoS Comput Biol*. 3; 13(3). 2017.
25. Abrams M A unified ethical frame for Big Data analysis. The Information Accountability Foundation big data ethics initiative. Available from <http://informationaccountability.org/publications/a-unified-ethical-frame-for-big-data-analysis>.
26. Nass SJ, Levit LA, Gostin LO, editors. *Beyond the HIPAA privacy rule: enhancing privacy, improving health through research*. Washington, D.C.: National Academies Press 2009.
27. Nelson GS. Practical implications of sharing data: a primer on data privacy, anonymization, and de-identification. SAS Global Forum Proceeding Paper 1884-2015, ThotWave Technologies, Chapel Hill 2015 (<https://support.sas.com/resources/papers/proceedings15/1884-2015.pdf>.)
28. Centers for Disease Control and Prevention. National Center for Health Statistics. Research Data Center, Sierra Leone Ebola Database website (<https://www.cdc.gov/rdc/b1datatype/sled.htm>.)
29. O'Connell ME, Plewes TJ. Sharing Research Data to Improve Public Health in Africa: A Workshop Summary National Academies of Sciences, Engineering, and Medicine. Committee on Population, Division of Behavioral and Social Sciences and Education, 2015 Washington, DC: The National Academies Press.

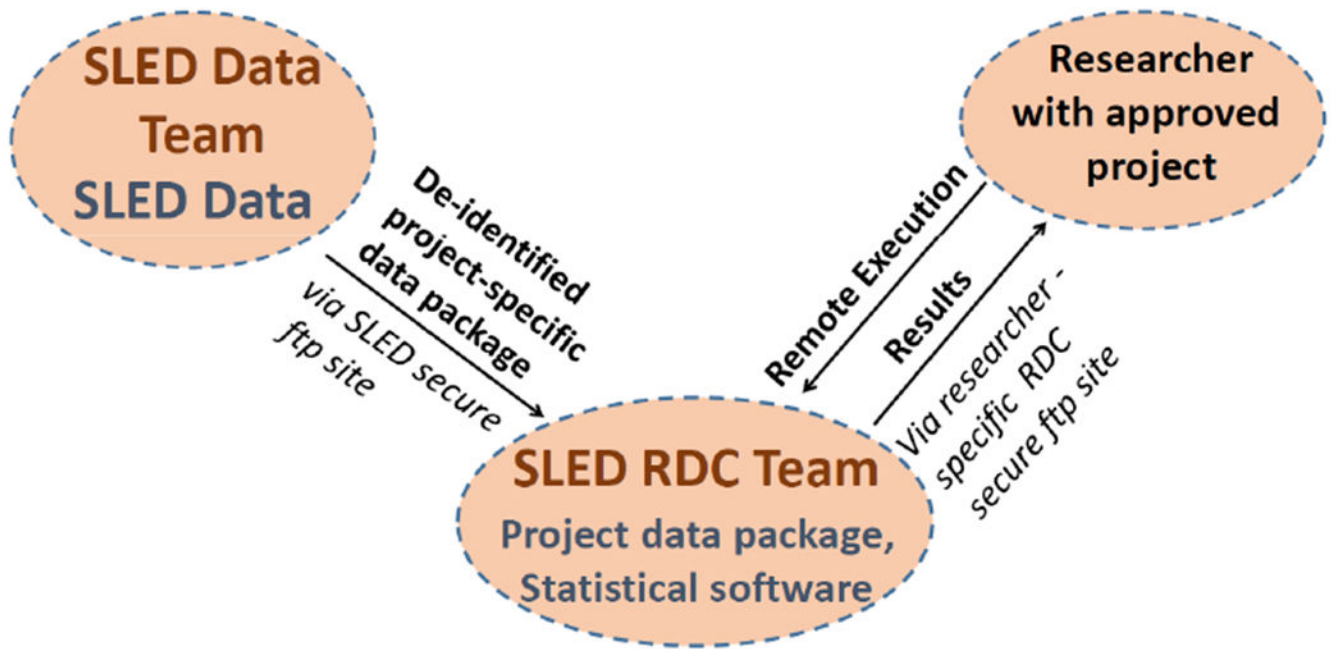


Figure 1. Research Programs Remote Execution using Sierra Leone Ebola Database Research Data Center (SLED RDC) Technology

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

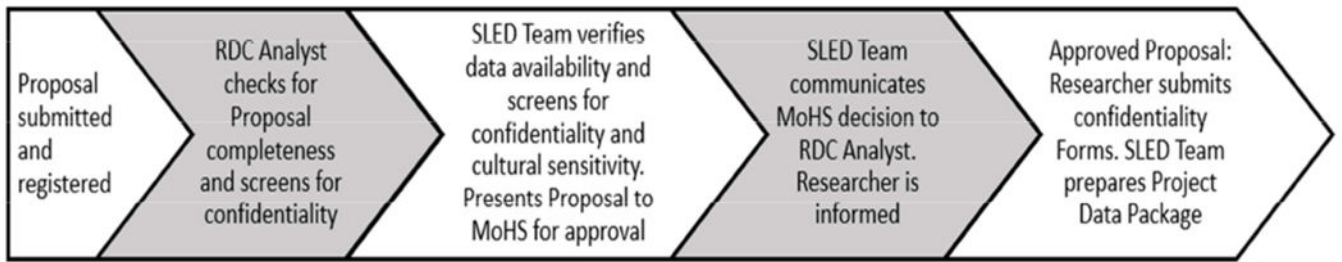


Figure 2.
SLED Research Proposal Approval Timeline

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript