Building a Centre of Excellence in Data Management in East Africa

Betty Agwang (✉ betagwang@gmail.com)

Makerere University Infectious Diseases Institute <https://orcid.org/0000-0002-4909-5595>

Yuka Manabe

Johns Hopkins University School of Medicine

Research

Keywords: Data management System, High quality data, Centre of excellence, Collaborations, East Africa

Posted Date: August 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-779411/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Building a Centre of Excellence in Data Management in East Africa

^{3,1} Betty Agwang, ¹Mariam Lule Namaweje, ^{2,4}Steven Reynolds, ², ¹Mark Ssennono, ¹Charles Luswata, ¹Perry Okoto, ¹Paul Kagwa, ¹Barbara Castelnovo, ¹Agnes Kiragga, ¹Henry Onen, ¹William Kasozi, ³Kevin Newell, ³Michael Duvenhage, ³Chris Whalen, ³Michael Tartakovsky, and ²Yukari C Manabe

¹Infectious Diseases Institute, Makerere University College of Health Sciences, Kampala, Uganda

²Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

³Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD USA

⁴Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD USA

Corresponding Author: ²Yuka C Manabe

Michael Tartakovsky and Yukari C Manabe contributed equally to the work included in this manuscript.

Funding from the National Institute of Allergy and Infectious Diseases, Division of Intramural Research and the Office of the Director, the Infectious Diseases Institute of Kampala, Uganda, and the Johns Hopkins University.

Abstract

Background: In resource-limited settings, there is a paucity of high quality data management systems for clinical research. The result is that data are often managed in high-income countries disadvantaging researchers at sites where the data are collected. An institutional data management system to address the data collection concerns of the collaborators and sponsors is a key institutional capacity element for high quality research. Our goal was to build a local data management center to streamline data collection and validation compliant with international regulatory bodies.

Methods: Leveraging established collaborations between Office of Cyber Infrastructure and Computational Biology of the National Institutes of Health and the John Hopkins University School of Medicine in the United States, the Infectious Diseases Institute at Makerere University built a data management coordinating center. This included mentorship from the NIAID International Centers for Excellence in Research and training of key personnel in South Africa at a functioning data center. The number of studies, case report forms processed and the number of publications emanating from studies using the data management unit since its inception were tabulated.

Results: The Infectious Diseases Institute data management core began processing data in 2009 with 3 personnel, hardware (network-enabled scanners, desktops, server held in Bethesda with nightly back up) and software licenses, in addition to on-site support from the NIH. In the last 10 years, 850,869 pages of data have been processed from 60 studies in Uganda, across sub-Saharan Africa, Asia and South America. Real-time data cleaning and data analysis occur routinely and enhance clinical research quality; a total of 212 publications from IDI investigators have been

published over the past 10 years. Apart from the server back-up services provided by the NIH, the center is now self-sustaining from fees charged to individual studies.

Conclusion: Collaborative partnership among research institutions enabled the IDI to build a core data management and coordination center to support clinical studies, build institutional research capacity, and to advance data quality and integrity for the investigators and sponsors.

Key word: Data management System, High quality data, Centre of excellence, Collaborations, East Africa.

Biomedical, public health, and health services research in developing countries is generating increasing volume and complexity of data as the number of research studies in sub-Saharan Africa (SSA) continues to grow [1]. Data management is a critical aspect of high quality research. The researchers, software analysts and developers, and biostatisticians need to collaborate using a data management system that addresses each stakeholder's needs including accurate data in real time, easily exportable data to create aggregate regular data reports, simple data integration, data dictionaries and edit check validations [1]. Good data management is not a goal in itself, but rather a key component to high quality research [5].

Importantly, data is often managed by the developed country partners and held outside the country where the data is generated [2]. This data ownership arrangement necessarily disadvantages local, indigenous researchers in the analysis and publication of data. The development of new user friendly data management systems and technology has allowed many countries to manage their own data within country. With the help of research capacity building programs, this has enabled indigenous researchers to own and analyze the data which, in turn, has led to sustainable capacity development. When data management is handled in the country where the data is

collected, local investigators develop expertise through forced interactions with data managers. Greater involvement with databases also increases insight into efficient and pragmatic data collection and, subsequently, improved protocol development skills.

The ultimate goal is to provide high-quality reliable data that is collected in accordance with the protocol and that requires less post-hoc cleaning. In the past, in-country data management was often handled using double data entry as the quality control process. A large amount of tedious direct data entry was often relegated to relatively poorly paid personnel without advanced degrees. Even with meticulous quality control procedures, there was a high likelihood of errors that had to be adjudicated. Study staff were often forced to do data entry due to the consistent backlog instead of concentrating on clinical work. With increasing data volumes, query resolution and auditing data entry and quality became more challenging; auditing data trails is very important because it monitors modifiable factors such as human error, and allows compliance with regulatory bodies. Manual entry led to long delays between patient assessments at the clinical sites and the availability of that information in a clean study database. In turn, accurate progress reports to manage implementation bottlenecks and corrective action were also delayed [6]. Real-time recognition of clinical trials attaining statistical stopping rules was a source of participant and institutional risk. Finally, clean dataset delays can result in repetitive case report form (CRF) errors that become more difficult to correct with time. These delays may also delay publication of study results [9]. We describe herein the adoption of an institutional data management system that addressed many of the concerns facing African research centers of excellence. [3]

Key features of a data management system appropriate for resource-limited settings

We established a data management unit that maintains high quality data and is able to meet stakeholder needs in a timely manner within a supportive institutional infrastructure at the IDI as a model for in-country data management in resource-limited settings. Attention was paid to cost since studies would pay for the service to ensure sustainability. This was accomplished by building data management skills in-house. With the growing number of multi-site remote clinical trials at the IDI, key features of an ideal data management system were:

- Simple technology and reliable infrastructure for data capture with option for paper CRFs and electronic transmission at convenience.
- Simple technology that does not require advanced hardware, minimal need for continuous electricity supply, and software for data transmission compatible with infrastructure at study sites.
- User-friendly data management system with semi-automated data entry.
Efficient study start-up through case report templates, database piloting, and standardized data operating procedures.
- Ability to create edit checks to increase data quality and automated weekly reports interpretable with limited training allowing regular data cleaning particularly for missingness.
- Following a needs assessment that helped to characterize scientists' research methods and data management practices, gaps and barriers were highlighted [6]. A decentralized data management system, ©DataFax/DFdiscover, was chosen where each site transmitted data through a fax machine and the data management center reviewed, verified, and validated data. ©DataFax/DFdiscover addressed many of IDI's logistical challenges commonly encountered in the management of large multi-center clinical trials and in resource-limited

settings (e.g., intermittent electricity). The system also supports Intelligent Character Recognition (ICR) and computerized receipt, logging and filing of data forms. After several layers of record review occur. After data collection is completed in a given study, database lock is performed following in-depth data quality checking [4]. Locking or closing a database is a crucial practice that is necessary in order to prevent unauthorized or unintentional changes after the final data entry, cleaning and analysis is complete.

Training of local teams to handle data

In July, 2009, a data manager and a data reviewer under the mentorship of a clinical trials manager from the NIAID International Centers for Excellence in Research (ICER) of the National Institutes of Health (NIH) were sent to South Africa for training in an existing data management unit using the same system (©DataFax/DFdiscover). The ultimate goal was to build a local data management center cognizant of the local research needs and designed to be integrating into the research context and the local health systems [9]. The first two studies using the data management unit started were an autopsy study describing the causes of death among HIV-positive patients in Kampala, Uganda [8] and a clinical trial for Multivitamins, HAART and HIV/AIDS in Uganda (TOV) Sites in Mulago.

Figure 1 : The Data Management Centre Structure and Data Flow

Some of the advantages include: reduced data processing time through timely collection of CRFs, improved quality control by use of quality control notes to flag and track CRF problems, improved two-way communication link between the study data management center and the participating clinical sites, CRF imaging, storage and retrieval system, simple technology at the clinical sites, study management tools and reports, controlled access to study data & CRFs, automated work flow management and increased efficiency through intelligent character recognition, thereby reducing personnel costs.

The cost of the IT infrastructure at the server end was estimated at USD \$102,650 at the beginning of the project in 2008. However, through collaborations with OCICB at the NIH the cost was reduced by using existing servers outside the country in Bethesda, Maryland. NIH also provided nightly back-up service which assured both data security and storage. Figure 1 shows the data flow and data management architecture at the center.

On the management center side, computers with very rapid processing speeds and large RAM for quick processing of data images (processor Intel dual core or higher, 1.0 GHz or higher, 1GB RAM or higher) are required. There are no restrictions to laptops or desktops, although Thin Clients have also proved to be very good option due to reduced power consumption. High-speed and reliable internet connections at the processing center are critical and mandatory.

The key personnel needed at the management center include, Team Lead, who directly supervises personnel at the center, an accountability officer who also has data management responsibilities, Data Managers who are cross-trained as programmers, Data Reviewers and Medical Coders. An IT technician was also needed especially with the set-up of equipment at the study sites. The human resource structure must address process flow since poor management will affect sites and communication with sites. Because there is direct contact between data personnel and Principal Investigators (PIs), this translates to better quality protocols and improved quality of data.

ASSESSMENT OF THE DATA MANAGEMENT CENTER

Figure 2: Total number of studies set up over the past 10 years

Figure 3: International PI's versus Local PIs

Figure 4: Countries with study sites in the data management center

Figure 5: CRF Volumes over time

Figure 6: Total number of papers published at the IDI

Figure 2 shows the total number of studies that have been set up in the database from the first training in South Africa through 2019. The number of studies has accelerated with 60 studies set up from 2014-2019. Figure 3 shows the numbers of local PIs versus the number of International PIs whose data has been managed by the data management Centre. Figure 4 shows the worldwide distribution of study sites managed by the data management Centre. Figure 5 shows the number of CRF pages submitted over the past 10 years. A total of 850,869 records were in the database as of December 2019.

The data management center in close collaboration with the research department and statistical department produces detailed validation reports, descriptive statistics, and plots for the investigators [10]. Access to the data and real-time data cleaning and database completion has led to improved publication productivity as shown in Figure 6.

CHALLENGES AND KEY LESSONS

All the study sites are trained on the data collection process focusing mainly on CRF completion. A key challenge to paper-based CRFs is that any change to a form creates difficulty in ICR data capture, review and analysis [3]. Although the site teams are encouraged to limit changes to the final versions of the CRFs, changes are still made and have resulted in database errors. Other data handling challenges included: sustainability due to high staff turnover due to repetitive task and high volume, shortage of well-trained biomedical research methodologists with data management skills in sub-Saharan Africa [7], delays in sending out and receiving quality control reports as volume of studies increases, vendor supplied software with yearly licensing costs, and data security and back-up of data to protect against catastrophic loss.

RECOMMENDATIONS

A key component of research capacity building is investment in the systems that underpin successful and high quality research. We were able to implement and grow a highly successful data management center appropriate for our low-income country setting that was compliant with regulatory standards, scalable, and high quality meeting international regulatory requirements for randomized clinical trials. The current lack of investment in data management in SSA can be attributed to a general lack of awareness of the fundamental role of data management and biostatistics in the conduct of research [6]. Holding the data at the sites where the research is conducted can empower local researchers and provide preliminary data for new proposals, and most importantly build capacity for scientific leadership. Training data managers and biostatisticians in SSA should be a capacity building priority [7].

DECLARATIONS

Ethical Approval and consent to participate

Not Applicable.

Adherence to national and international regulations

Not applicable.

Consent of publication

Not applicable.

Availability of data and materials

Not Applicable.

Competing interests

The authors declare that they have no competing interests

Funding

Not applicable.

Authors' Contributions

Y.C.M and M.T conceived of the presented idea, B.A, M.L.N drafted the manuscript. S.R, M.S, W.K, C.L, P.O, P.K, B.C contributed information to the manuscript. Y.C.M, A.K, H.O, K.N, M.D, C.W, M.T, B.C performed critical revisions of the manuscript. And all authors approved the final manuscript

Acknowledgements

We are indebted to the Office of Cyber Infrastructure and Computational Biology at the National Institute of Health and Johns Hopkins University School of Medicine for sustained essential support and their vision in creating the IDI Data Management Center. We would also like to acknowledge James D. Campbell for his key role in highlighting the need for a data management system at the IDI.

.

REFERENCES

1. Stephen B. Johnson, Frank J. Farach, Kevin Pelphey, Leon Rozenblit: **Data management in clinical research: Synthesizing stakeholder perspectives**. J. Biomed. Inform. 60 (2016) 286 -293.
2. Nicholas R. Anderson, E. Sally Lee, J. Scott Brockenbrough, et al: **Issues in Biomedical Research Data Management and Analysis: Needs and Barriers**. J Am Med Inform Assoc. 2007;14:478 - 488.
3. Yukari C Manabe, Elly Katabira, Richard L Brough, Alex G Coutinho, Nelson Sewankambo, Concepta Merry: **Developing independent investigators for clinical research relevant for Africa**. Health Research Policy and Systems 2011, 9:44.
4. Qazi Sadeq-ur Rahman, Mohammad Shahidul Islam, Belal Hossain, Tanvir Hossain et al: **Centralized Data Management in a Multicountry**, Multisite Population-based Study. Pediatrics Infect Dis J 2016;35: S23–S28.
5. Mark D. Wilkinson et al: **The FAIR Guiding Principles for scientific data management and stewardship**. Scientific Data 3:160018. DOI: 10.1038/sdata.2016.18.
6. Misrak Gezmu, Victor DeGruttola, Dennis Dixon, Max Essex, Elizabeth Halloran, Joseph Hogan et al: **Strengthening Biostatistics Resources in Sub-Saharan Africa: Research Collaborations through U.S. partnerships**. Stat Med. 2011 March 30; 30(7): 695–708.
7. Rhoderick Machekano, Taryn Young, Simbarashe Rusakaniko, Patrick Musonda, et al. **The Africa Center for Biostatistical Excellence: a proposal for enhancing biostatistics capacity for sub-Saharan Africa**. Statist. Med. 2015, 34 3481–3489.

8. Alex C. Ezech, Chimaraoke O. Izugbara, Caroline W. Kabiru et al: **Building capacity for public and population health research in Africa: the consortium for advanced research training in Africa (CARTA) model.** Global Health Action 2010, 3: 5693.
9. Gini Rosa, Schuemie Martijn, Brown Jeffrey, et al: **Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison Among EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies.** eGEMs (Generating Evidence & Methods to improve patient outcomes), Vol. 4 [2016], Iss. 1, Art. 2.
10. Adelheid W. Onyango, Alain J. Panel, and Mercedes de Onis: **Managing data for a multi-country longitudinal study: Experience from the WHO Multicentre Growth Reference Study.** Food and Nutrition Bulletin, 2004 vol. 25, no. 1 (supplement 1). The United Nations University
11. Wafaie Fawzi, David Guwatudde, Yukari Manabe: **MULTIVITAMINS, HAART AND HIV/AIDS IN UGANDA:** Medicine, 2016 - ncbi.nlm.nih.gov

List of abbreviations

IDI- Infectious Diseases Institute

NIH- National Institute of Health

NIAID- National Institute of Allergy and Infectious Diseases

OCICB- Office of Cyber Infrastructure and Computational Biology

HIV- The Human Immune Deficiency Virus

HAART- Highly Active Antiretroviral Therapy

CRF- Case Report Form