

GLOBAL PROGRAMS for Research & Training

Affiliate of the University of California, San Francisco

VACANCY ANNOUNCEMENT

HISTAC Data Manager

Nairobi, Kenya

Closing Date – 7 December 2021

BACKGROUND

Established in 2016, the Global Programs for Research and Training, based in Nairobi, is a Non-Governmental Organization registered in Kenya to represent the University of California, San Francisco (UCSF) East Africa projects. UCSF's Global Strategic Information (GSI) has worked closely with CDC/PEPFAR for nearly 10 years with a focus on strategic information (SI) and development of associated Health Information Systems (HIS). Additionally, GSI has more than 10 years' experience working on HIS in over 15 countries in sub-Saharan Africa (SSA), the Caribbean, Southeast Asia, and eastern Europe, providing high-level strategic thinking, technical assistance (TA) in the development of all levels of HIS systems, data presentation and interpretation, and local capacity building. UCSF has always worked closely with multiple stakeholders on the ground including MOHs, institutions of higher education, and implementing partners (IP). Currently, UCSF's GSI will lead a Health Information Systems Technical Assistance Consortium (HISTAC) with a focus on an Electronic Medical Record (EMR) OpenMRS HIV Reference Implementation (OHRI) and Data Integration Strategies and Implementation (DISI) (e.g., National Data Repository, National Data Warehouse, Health Information Exchange, Patient Management Solution, etc.). The project will span over multiple countries such as: Uganda, Kenya, Ethiopia, Haiti, and Nigeria. Additionally, UCSF will be providing technical assistance on data quality to Cambodia as part of this same award mechanism.

The HIS Data Manager will primarily be responsible for generating and maintaining several synthetic datasets to be used for comprehensive testing of the newly developed OpenMRS modules: HTS, Care & Treatment, PMTCT and COVID case management. S/he will also be providing data management and insight for reporting (including PEPFAR reporting) and visualization requirements for these systems. Related to this, the DM will also define minimum datasets for national program monitoring and patient care coordination within programmatic health responses scope of work, such as HIV case-based surveillance. Under the Senior Business Analyst, s/he will work closely with UCSF HQ, University of Nairobi, HIS Developers & Business Analysts, and other programmatic staff and stakeholders.

Specifically, data management encompasses the following dimensions of which the candidate applying should have adequate knowledge and demonstrated application:

- **Analysis and visualization:** Approved methods for cleaning and transforming data to enable analysis, visualization, and use at various levels (MOH, stakeholders, etc.)
- **Data management:** Collection, management, integration, analysis, visualizing and reporting, as well as supporting data use and interpretation of data in line with proper data management procedures and best practices using the SQL language. Synthetic data generation using appropriate techniques and platforms (such as R or other)
- **Appropriate use:** Ensuring privacy and confidentiality through the definition and construction of minimum use data sets and guidance on appropriate statistical practices that achieves the objectives of an internal/external data user
- **Data security:** Verifying data users as well as the implementation and documentation of appropriate administrative, technical, and physical safeguards for all data (along with appropriate penalties)
- **Data quality:** Ensuring data standardization, verification, and validation procedures are implemented and documented to ensure data accuracy, availability, timeliness, completeness, credibility, and solid understanding of its limitations
- **Public disclosure:** Documenting all public data requests and ensuring appropriate communication between data stewards and data users for arranging data disclosure prioritization and procedures

Roles and Responsibilities:

Analysis and Visualization

- Incorporating technical and programmatic specifications to generate synthetic datasets for testing newly developed OpenMRS forms using R
- Cleaning and ETL of data using SQL scripts (stored procedures, functions, etc.) periodically
- Defining minimum datasets needed for national/subnational level program monitoring (ie MER indicators) and patient care coordination (i.e. continuity of care between facilities)
- Integrating country-specific data for data quality, validation, de-duplication, and proper linkage.
- Inspecting externally created SQL scripts for above data management for accuracy
- Use of visualization tools such as Jasper Reports to display data on graphs, maps, and line-lists

Documentation

- Creation of documentation and a solid understanding of all data flows (from source to analysis and use) for any data as required
- Maintenance and understanding of all SQL-based documentation which details data cleaning, standardization, validation, and analytical procedures
- Maintenance and ongoing analysis of data quality as well as tracking of all 'bugs'
- Documenting and cataloging all data elements (data dictionaries and metadata) for the various datasets as well as the relationships between them (e.g., ERD's for data linkages)
- Documenting all changes to variables and data structures (i.e., change management)
- Documenting minimal use dataset creation criteria and parameters

Communication and Coordination

- Coordinating with the informatics team to ensure proper prioritization of minimal dataset development and fixing data quality issues

Technical Assistance

- Ensuring appropriate use through guiding analytical approaches and ensuring that correct conclusions are drawn from source data
- Providing guidance on data visualizations and reporting for national/subnational use cases
- Creation of minimal use datasets to meet external data user needs, requiring an understanding of SQL and procedures/functions

Technical Documentation

- Ensuring that technical guidelines for anonymization of datasets are followed for external data use while still maintaining data linkages across appropriate datasets

Minimum Requirements:

- Master's degree in Data Science, Data Management, Computer Science, Software Engineering, or a related field with sufficient training in database related disciplines
- At least five years of working experience managing active large study/surveillance datasets
- Fluent in R, must be able to demonstrate this experience
- Demonstrable experience with at least one of the following: Oracle, SQL Server, MySQL, PostgreSQL
- Fluent in complex data manipulation using SQL, especially more advanced scenarios
- Demonstrable advanced expertise in writing complex SQL queries and statements.
- Experience working in the health sector, public health, and/or PEPFAR programs
- Demonstrated good communication and interpersonal skills, especially working remotely and over multiple time zones
- Ability to coordinate with multiple partners in various countries to carry out duties as assigned, especially working remotely

Other Desired Skills:

- Certification in one or more popular database management systems (DBA in Oracle, SQL Server or similar)
- Demonstrable experience with data visualization using at least one of the following platforms: PowerBI, Tableau, Qlickview
- Background in statistics and/or data analysis
- Experience with international health/research/data projects

TO APPLY

Qualified Kenyan nationals are encouraged to apply.

All applicants must address each selection criterion detailed in the **minimum requirements** above with specific and comprehensive information supporting each item. All applications must include the following:

- Cover letter
- Current CV with names and telephone numbers for three referees

Applications should be sent by email to HR.Kenya@ucglobalprograms.org with the email subject, **Data Manager** by **Tuesday, December 7, 2021**

Only short-listed candidates will be contacted.