Data Scout: A Metadata Navigator & Streamlining Tool for Countries

Submitted by Hannah Cooper (Cooper/Smith) on January 19, 2018 - 10:01am
Last revised by Web Producer on June 21, 2018 - 3:09pm.
Proposal Status: In Review

Executive Summary

The Government of Malawi (GOM) is in the process of harmonizing and consolidating data collected across the health sector. The number of data collection tools is unknown and a comprehensive list of data elements recorded not available. At facility level, health workers (HWs) must record patient level transactions and aggregate data in Government registers and reports, as well as those required by donors. The sheer volume of information that HWs must manage and report leads to poor data quality, sub-optimal reporting rates, inefficient processes, and excessive burden for an already stretched health workforce. Additionally, managers and policy makers struggle to identify the right data and reconcile discrepancies across systems and reporting streams, leading to insufficient data use. Supporting document A highlights the flow of data for the HIV program alone.

The Kuunika Project is a four-year effort, funded by the Bill & Melinda Gates Foundation, to catalyse the integration, use, and governance of health data in Malawi. Through Kuunika, Cooper/Smith is supporting the GOM and technical partners to identify, harmonize, and streamline health data. The ultimate goals are to reduce collection burden, digitize records, improve access, improve data quality, and foster a culture of continuous data use.

Cooper/Smith is supporting the Ministry of Health's Central Monitoring and Evaluation Division (CMED) with accurately documenting data elements and mapping common elements across systems and paper tools. Cooper/Smith is currently developing a user-friendly digital application: "**Data Scout**" (previously called "Metadata Navigator") that allows users to author, visualize, and manage metadata across the entire health sector. (Supporting Document B outlines the high-level framework for categorization). In late 2017, the vision for this application was approved by CMED and currently being incorporated into the MOH's latest Monitoring & Evaluation Framework. Cooper/Smith is looking for funding from Digital Square to turn Data Scout into a global good that can be scaled for broader application in multiple countries while remaining tailored to meet a country's specific requirements.

Consortium Team

The team consists of Cooper/Smith and CMED collaborators. We are looking for additional collaborators both on developing the Data Scout user interface (UI) and to assist with developing the navigator into a scaleable and tailored tool.

Cooper/Smith is a technical assistance organization that uses hard data to increase the effectiveness & efficiency of development programs worldwide. We have experience leading full-cycle, large-scale data initiatives including requirements collection, defining metrics, tool and system development, piloting and scale-up, and applied used of results. We have deep knowledge of strategic planning, evidenced-based resource allocation, and DHIS2 roll-out and training (specifically DATIM – PEPFAR instance of DHIS2). Our staff have served in many roles: researchers, economists, strategic information specialists, policy advisers, software developers, and clinical coordinators.

Cooper/Smith supports large-scale data systems projects in southern Africa funded by the Bill & Melinda Gates Foundation, the Digital Impact Alliance (DIAL), and the Center for Infectious Disease Research Zambia (CIDRZ); advises the Government of Canada's Minister of International Development, the End AIDS Coalition, the Center for Strategic and International Studies (CSIS) and the International Fund for Agricultural Development; as well as participates in numerous global and local technical and advisory groups.

Hannah Cooper, Co-Founder and Managing Director of Cooper/Smith will serve as the organizational management lead.

Project Description

Cooper/Smith has catalogued and stored 5000+ data elements in a relational database. We are in the process of selecting a software company to

develop the UI. In tandem, should this proposal receive funding from Digital Square, we will work to develop out Data Scout for broader applicability while ensuring that it still has the necessary functionality to tailor it to specific country level requirements. Below are the principles and high-level requirements guiding Malawi's Data Scout development:

Principles

- The application provides an intuitive and seamless **user experience** for use by anyone ranging from frontline health workers to program managers, district health officers to MOH staff
- The design process is **agile and user-centered**
- The application makes use of **open-source** software and technologies as much as possible
- The application **integrates** with other systems in the digital landscape/architecture of Malawi and other low-and-middle-income countries, in particular the terminology service (TS) and interoperability layer (IOL), as defined by the OpenHIE framework
- The application is being designed in an **extensible** manner to accommodate future feature requests or design changes

High-level Requirements

User Interface

- The application is web-based and accessible through any modern web browser
- Data elements, indicators, data sources, and stakeholders are displayed as nodes connected by relationships in a **graph-based visualization**, akin to Linkurious
- The UI should support **interactive features**, including but not exclusive to zoom-in and zoom-out, dynamic rearrangement of nodes/relationships, etc.
- Each node and relationship is associated with predefined **attributes/properties** (aligned with CMED requirements) which can be edited in a side panel
- Similar or duplicated data elements or indicators are visually highlighted or grouped to enable seamless **deduplication**

Search

- **Search** is enabled across all content, with the ability to perform fuzzy search and faceted search
- Search results are **sortable and filterable** as directed by the user

Database

- All data required for this application are stored in a database **conducive** to visualization and search requirements as described above, and any additional analytics required
- This may involve the use of a **graph database**, such as Neo4j, or another non-relational (NoSQL) database

Verification Workflow

- All submissions (additions/edits/deletions to data elements, indicators, data sources, and stakeholders) are subject to a **two-staged verification workflow** that involves the data custodian (stage 1) and CMED (stage 2)
- Each data element and indicator are assigned an organization acting as the data custodian (stakeholder), which assigns one or more representatives with the authority to **approve or reject** submissions for stage 1

User Management

- **Role-based access control** (RBAC) are instituted to define privileges, roles, and users within the application
- Roles group together **privileges**, which include viewing, editing, and approving
- Users' **roles** should be confined to the data elements and indicators under the custodianship of the organization to which they belong

Data Import/Export/Integration

- The pre-existing relational database are **imported** into a graph or non-relational database as required
- **Export** of data (search results or according to specified criteria) to CSV files is enabled
- Bi-directional integration with the **terminology service and interoperability layer** is accommodated – these may comprise of platforms such as Open Concept Lab and OpenHIM
- Bi-directional integration with the health management information system, namely **DHIS2**, should also be accommodated

Timeline

The Malawi prototype will be completed by the end of June 2018. The global Data Scout prototype will be available by end June 2019.

Supporting Documents: appendix_a_-_malawi_hiv_dataflow.pdf
appendix_b_highlevelcategorizationframework_datascoutmalawi.pdf