Columbia Global eHealth Program OpenMRS Concept Dictionary Concept Document

Andrew S. Kanter, MD MPH

Assistant Professor of Clinical Biomedical Informatics and Clinical Epidemiology at the Earth Institute, Columbia University
Director of Health Information Systems/Medical Informatics for the Millennium Villages Project

Edward Johnson, MA BEd CNI

Fellow in Operational Linguistics, Wolfson College, University of Cambridge
Director of Prolingua Ltd.

Table of Contents

Background

The Millennium Villages Project (MVP), based at the Earth Institute at Columbia University, is a new bottom-up approach to lift developing country villages out of the poverty trap that afflicts more than a billion people worldwide, and help them get on the path of self-sustaining development. Millennium Villages are either existing or being developed in 10 countries throughout Africa (there are 12 millennium village sites running in 10 countries in Sub-Saharan Africa).

To achieve this goal, the MVP requires an information and communication system which can provide both necessary infrastructure for monitoring and evaluation, and tools for communicating among villages, cities and countries at and between several clinical and administrative levels. This system is called the Millennium Global Village-Net (MGV-Net). It will take advantage of the latest in open source software (OpenMRS), databases (MySQL), interface terminology, centralized concept dictionary, and use appropriate technology locally for data entry.

The centralized concept dictionary requires a website for the management and collaborative generation of the fundamental elements required by the system's multi-lingual, multi-layered functionality. The site will present the terms and other data elements, descriptions, definitions and concepts used or destined for the system and will provide the medium for participation by the project's partners on terminological choices and relationships, translation, normalization and other inter-community semantic concerns.

Description of the Concept Dictionary

Format/Structure

The Columbia Global eHealth Program (CGeP) concept dictionary is based upon the OpenMRS concept architecture as implemented by the Millennium Villages Project among others. The structure of the dictionary is intimately tied to the OpenMRS system and therefore follows the OpenMRS data model. The current format of the tables corresponds to the 1.4.2.02 version of the database and is described below.
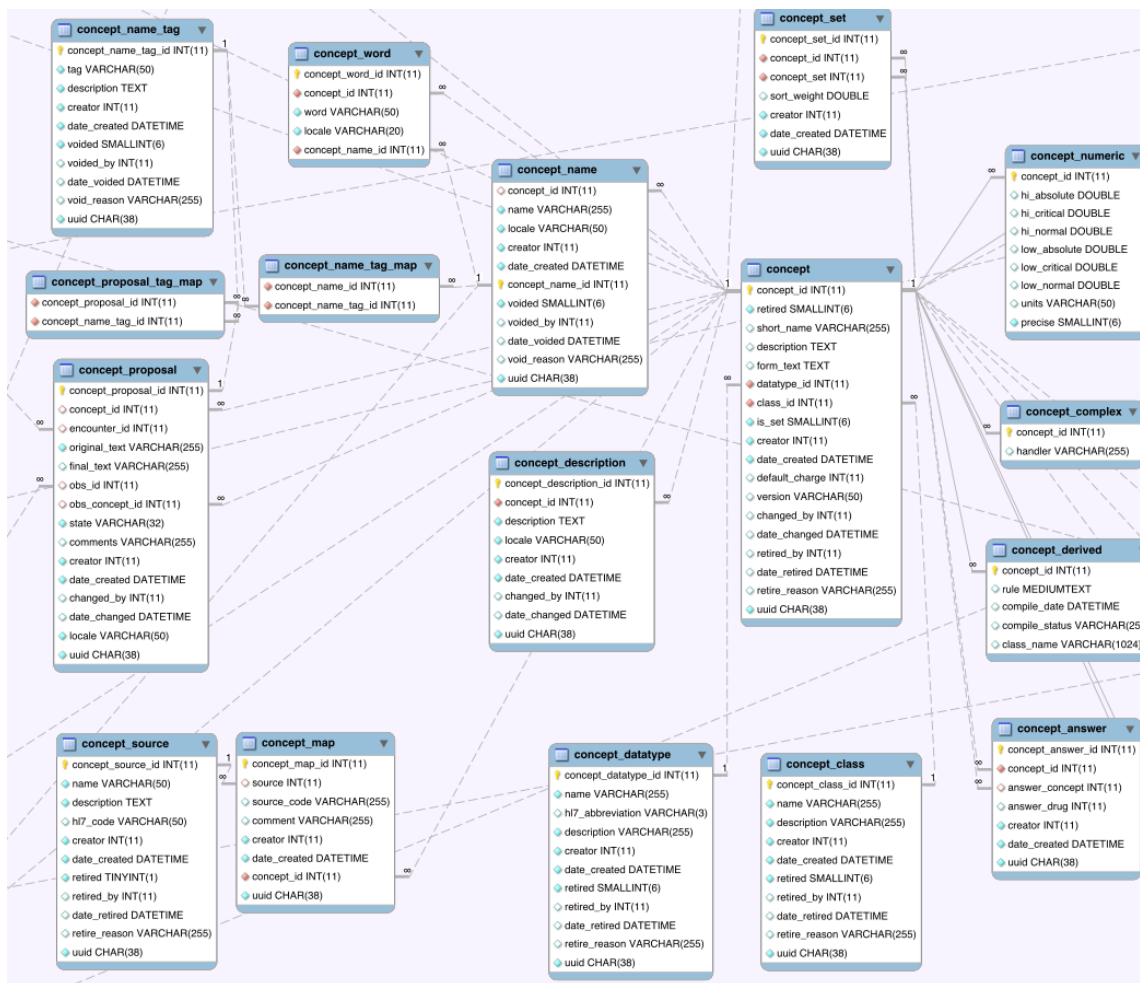
Scope

Although some degree of internal contextual marking is anticipated, the CGeP concept dictionary in its fullest sense is not disease-specific nor context-specific. The initial purpose of the concept dictionary is to support primary health care in Africa and the present work necessarily reflects the diseases and workflows expected in the rural health clinics of MVP. The dictionary is *not* however exclusive to that originating context. The harmonization and normalization that occurs through the collaboration of different organizations and projects makes it possible for the dictionary to grow to serve other functions, without depleting its utility for MVP.

The current dictionary contains the following major types of concepts:

1. Diseases/Diagnoses/Problems

2. Procedures/Tests

3. Medications

4. Survey Questions (birth/death registration, verbal autopsy, others)

5.   Antenatal Care

6.   AMPATH HIV concepts

Coding/Mapping

The primary purpose of the shared concept dictionary is to achieve the maximum practical degree of semantic interoperability. To ensure that this principle is maintained within and outside of MVP, as many concepts as possible are mapped to standard reference terminologies. Each concept type has a preferred mapping that allows for additional secondary maps if necessary. The primary reference maps for the types above are as follows:

1.   **Diseases/Diagnoses/Problems**
         (SNOMED CT, ICD-10-WHO, IMO Problem IT)

2.   **Procedures/Tests**
         (SNOMED CT, IMO Procedure IT, LOINC)

3.   **Medications**
         (RxNORM)

4.   **Survey Questions (birth/death registration, verbal autopsy, others)**
         (SNOMED CT, others)

Translations

The MVP implementation of OpenMRS extends to numerous countries in Africa and elsewhere. The user community is professionally multi-layered and linguistically diverse. This requires close attention to those aspects of the common data dictionary that will promote multi-community working. Part of the requirement is to achieve terminological stability, reduce ambiguity and develop strategies to deal with cultural variation. Also required is the ability to render in different languages as many of the concept names, textual descriptions or labels, and definitions within the dictionary as possible. Both aspects must be conducted in tandem by a specialized translation service in cooperation with the technical developers and user community. This process has been initiated within a section of the MGV-Net development known as the Terminology Service Bureau (TSB).

The MVP methodology is to connect the TSB to the MVP concept dictionary to facilitate these processes. Currently the TSB operates on a different platform than OpenMRS, but an MVP-supported coding effort will be developing the TSB as a module within OpenMRS allowing other institutions to create their own TSBs. In the MVP case, we expect that initially all translation–related activity will go through the MVP data dictionary. In the following figure, all concept proposals come to MVP and then are viewed within the TSB for potential translation. Those concept names (in different languages) are then passed back to the MVP OpenMRS system for distribution. In the future, it would be better to have the TSB directly connect to the OpenMRS Concept Collaborative (OCC) and therefore be able to apply translations to non-MVP-maintained concepts. Perhaps as part of an OpenMRS module, this would be easy.
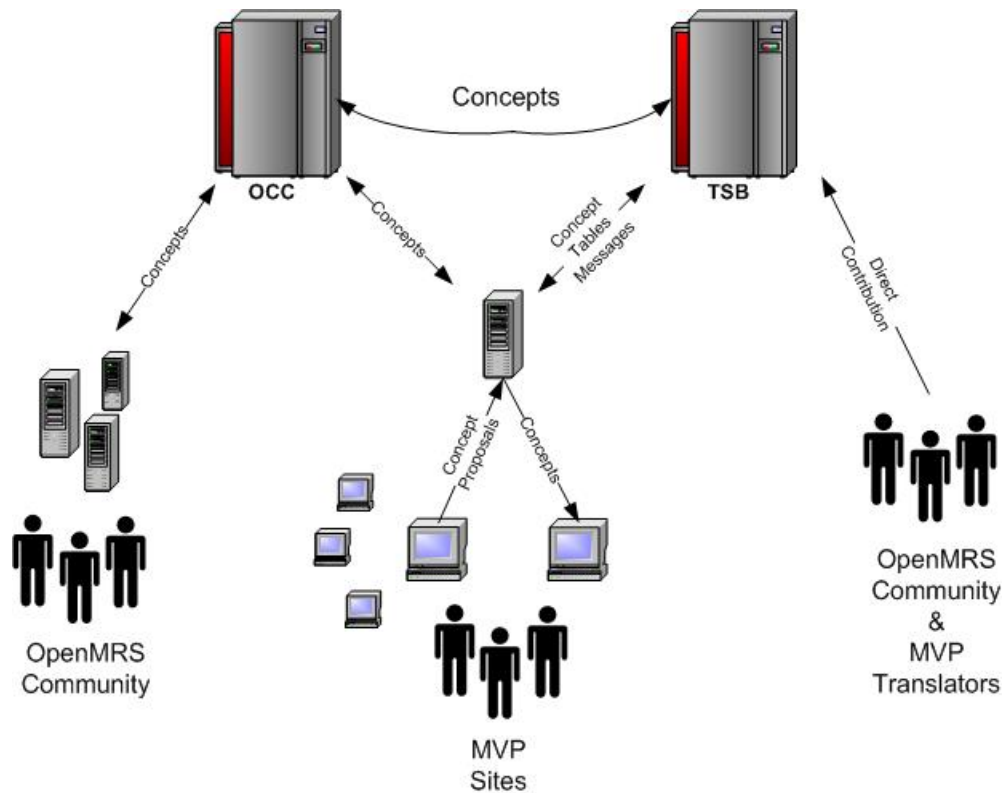
Figure 1: Possible relationships between the OpenMRS Concept Collaborative and Terminology Service Bureau

Creating Subsets

Currently, the MVP-maintained concept dictionary has tens of thousands of concepts encompassing a broad spectrum of concept domains as described above. This sets up an implementation problem for many sites. MVP would like to provide a methodology for creating specific subsets of concepts that could be used by sites under particular circumstances. The system might allow local OpenMRS implementations to use the subsets for the majority of functions, then access the full MVP-maintained dictionary for additional searches when necessary, and finally accessing the full OCC if other concepts are needed.

One method of only using a portion of the MVP-maintained dictionary would be to use the OCC to subscribe to specific elements of the MVP dictionary. However, at this point the connection to the OCC is theoretical only.

Requesting New Concepts

For this shared data dictionary to work, it is critical that new concepts can be added to the dictionary quickly, and that each user of the system be able to quickly get new concepts reviewed and eventually released into the dictionary. This process can occur through several methods. The easiest method would be simply via email request. This can occur one at a time, or with an attached spreadsheet. It is important that at least an initial effort to check whether the concepts already exist be performed by the requestor. Another method

is via OpenMRS itself. As part of a new Google Summer of Code project, as well as some additional coding, it should soon be possible for a subscriber to the concept dictionary, if a user in the OpenMRS system (remote), to request a new concept via OpenMRS. This would then filter up to the MVP (host) server and be processed.

MVP is also considering a more advanced tracking system for proposed concepts which would allow requestors to follow the request through the editorial, translation and approval process. MVP intends to make updates of the concept dictionary available as regularly as possible, particularly if needed by a project.

Requesting Mapping Changes

It is also likely that there will be queries or updates to cross-mapping within the concept dictionary. Currently, the OpenMRS reference maps are not well used, nor in broad implementation. However, the infrastructure for the mapping tables has been agreed to by the OpenMRS leadership. It needs to be built into the release tables by version 1.9. For now, the cross-maps to ICD-10-WHO, SNOMED CT, LOINC or RxNORM are held in the MySQL database.

Governance of the Dictionary

It would be natural for non-MVP users of the concept dictionary to be concerned about possible changes to the dictionary, formatting, or even release cycles. Although we are primarily required to support the CGeP projects, we are conscious of the need to include others in the overall planning, direction and support of the multi organization harmonization effort. Although individual editorial decisions cannot practically be made by consensus, we do intend to open up governance of the concept dictionary. An editorial board led by Edward Johnson at the University of Cambridge and Andrew Kanter at Columbia University will have primary responsibility for the content and the mappings. A second advisory board will be made up of representatives of the different groups using the dictionary, representatives from standards bodies including WHO, and other relevant representatives. The idea is that the advisory board would provide guidance and input to the overall direction and functions of the shared resource, but would not have day-to-day responsibilities.