CRVS Fellowship report
SmartVA and DHIS2 integration in Sri Lanka

March 2019

Resources available from the University of Melbourne, Bloomberg Philanthropies Data for Health Initiative

CRVS course prospectuses
These resources outline the context, training approach, course content and course objectives for the suite of CRVS trainings delivered through the Bloomberg Philanthropies Data for Health Initiative. Each course focuses on a specific CRVS intervention or concept, and is designed to support countries to strengthen their CRVS systems and data.

CRVS Fellowship reports and profiles
The CRVS Fellowship Program aims to build technical capacity in both individuals and institutions to enhance the quality, sustainability and health policy utility of CRVS systems in Fellows' home countries. *Fellowship reports* are written by Fellows as a component of the program, and document, in detail, the research outcomes of their Fellowship. *Fellowship profiles* provide a summary of Fellows' country context in relation to CRVS, an overview of the Fellowship experiences, the research topic and the projected impact of findings.

CRVS analyses and evaluations
These analytical and evaluative resources, generated through the Initiative, form a concise and accessible knowledge-base of outcomes and lessons learnt from CRVS initiatives and interventions. They report on works in progress, particularly for large or complex technical initiatives, and on specific components of projects that may be of more immediate relevance to stakeholders. These resources have a strong empirical focus, and are intended to provide evidence to assist planning and monitoring of in-country CRVS technical initiatives and other projects

CRVS best-practice and advocacy
Generated through the Initiative, CRVS best-practice and advocacy resources are based on a combination of technical knowledge, country experiences and scientific literature. These resources are intended to stimulate debate and ideas for in-country CRVS policy, planning, and capacity building, and promote the adoption of best-practice to strengthen CRVS systems worldwide.

CRVS country reports
CRVS country reports describe the capacity-building experiences and successes of strengthening CRVS systems in partner countries. These resources describe the state of CRVS systems-improvement and lessons learnt, and provide a baseline for comparison over time and between countries.

CRVS technical guides
Specific, technical and instructive resources in the form of *quick reference guides, user guides* and *action guides*. These guides provide a succinct overview and/or instructions for the implementation or operation of a specific CRVS-related intervention or tool.

CRVS tools
Interactive and practical resources designed to influence and align CRVS processes with established international or best-practice standards. These resources, which are used extensively in the Initiative's training courses, aim to change practice and ensure countries benefit from such changes by developing critical CRVS capacity among technical officers and ministries.

Author

Dr Chamika H Senanayake, Ministry of Health, Nutrition and Indigenous Medicine, Sri Lanka.

Suggested citation

Senanayake, CH. *Fellowship report: Smart VA and DHIS2 integration in Sri Lanka*. CRVS Fellowship reports and profiles. Melbourne, Australia: Bloomberg Philanthropies Data for Health Initiative, Civil Registration and Vital Statistics Improvement, University of Melbourne; 2019.
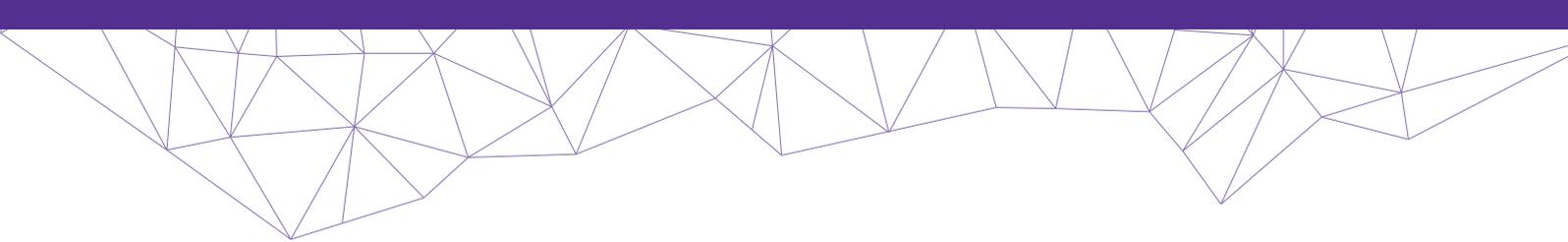
Contents

Acknowledgements

Fellowship report: SmartVA and DHIS2 integration in Sri Lanka

Background

Verbal autopsy (VA) is a tool for ascertaining the cause of death where medical certification of cause of death is absent.[1] As part of the Bloomberg Philanthropies Data for Health (D4H) Initiative, in February 2016, Sri Lanka began an extensive pilot study of SmartVA implementation in three districts. As part of the pilot, VA data were collected using the Open Data Kit (ODK) Collect tool via mobile devices, which were then uploaded to an ODK Aggregate server. Raw data were downloaded manually from the server in batches into Comma Separated Values (CSV) files. The respective CSV files were run through the Institute for Health Metrics and Evaluation's (IHME) SmartVA tool, which applies the Tariff 2.0 algorithm to generate CSMFs.

After successful completion of the pilot study, several key problems were identified including:

> VA was piloted in three districts to help identify and resolve problems before wider implementation.

- A lack of a platform to monitor collected results

- Difficulties in having to perform numerous manual processes using varied software packages to obtain final Cause Specific Mortality Fraction (CSMF) graphs

- The static nature of VA results-based data visualization

- The lack of data interoperability between SmartVA and other systems.

These problems are not unique to Sri Lanka, but highlight the challenges faced when countries attempt to integrate VA into civil registration and vital statistics (CRVS) systems,[2] which has been a major work area in the D4H Initiative.[3]

> Sri Lanka is one of the first countries to test the integration of data from VA into their routine health information system (DHIS2).

In order to address these problems, the University of Melbourne, working in partnership with BAO systems[4] initiated the development of a SmartVA-DHIS2 integration module to streamline the manual processes into a single automated process. Currently, Sri Lanka is one of the first countries to commence adaptive testing of this solution.

The District Health Information System, second version (DHIS2) is a free and open-source tool for the collection, validation, analysis and presentation of aggregate and patient-based statistical data, tailored (but not limited) to integrated health information management activities.[5] DHIS2 is developed by the Health Information Software Program (HISP), which is an open and globally distributed process with developers worldwide and coordinated by University of Oslo (UiO) with support from NORAD and other donors.[5] Currently DHIS2 software is used in more than 67 lower and middle-income countries.[6] Sri Lanka has adopted DHIS2 as the main health information system (HIS) within many departments of the Ministry of Health, Nutrition and Indigenous Medicine (MoHNIM).

1   WHO Verbal autopsy standards: ascertaining and attributing causes of death [Internet]. WHO. Available from: http://www.who.int/healthinfo/statistics/verbalautopsystandards/en/

2   de Savigny D, et al. Integrating community-based verbal autopsy into civil registration and vital statistics (CRVS): system-level considerations. *Global Health Action* 2017 (10):1272882

3   Challenges associated with automated VA training and rollout. CRVS development series. Melbourne, Australia: Civil Registration and Vital Statistics Improvement, University of Melbourne, Bloomberg Philanthropies Data for Health Initiative; 2018.

4   BAO Systems – DHIS 2 Support and Hosting [Internet]. Available from: https://baosystems.com/

5   DHIS2 User Manual [Internet]. Available from: https://docs.dhis2.org/master/en/user/dhis2_user_manual_en.pdf

6   DHIS 2 In Action | DHIS2 [Internet]. Available from: https://www.dhis2.org/inaction

Methods

Developing a SmartVA-DHIS2 integration module for Sri Lanka was achieved in five main steps:

1. Setting up DHIS2 instance and stability testing
2. Analysing of org-unit hierarchies
3. Importing VA metadata into DHIS2
4. Configuring the script
5. Localising the ODK Collect form.

Testing DHIS2 instance set-up and stability

It is important to test the behaviour of the linkage script in different versions of DHIS2 under different Linux Operating Systems (OS) before establishing country-specific DHIS2 instances. This can be done using server virtualization software such as Tomcat, WAMP, and MAMP for Windows and Macintosh systems.

For testing purposes, different Linux OS were incorporated into virtual environments. Four test cases were performed using three versions of DHIS2 instances on three Linux operating systems, as summarised in **Table 1**.

Table 1 Summary of testing

Test No #	Test 1	Test 2	Test 3	Test 4
Linux Distribution	Ubuntu Desktop	Ubuntu Desktop	Ubuntu Desktop	CentOS
Version	16LTS	16LTS	18LTS	7
Postgresql version	9.5.13	9.5.13	10.4	9.5.13
Java Version	JDK 1.8.0_171	JDK 1.8.0_171	Java-8-oracle	JDK 1.8.0_171
Apache-Tomcat version	7	7	8	7
DHIS2 version	2.28	2.29	2.30	2.29

Analysing org-unit hierarchies

The organisation (org)-unit defines the organisation hierarchy of the particular DHIS2 instance, such as how the health facilities, administrative areas and other geographical areas are arranged with respect to each other.[7] A thorough analysis of the current org-unit hierarchy and its relationship with health boundaries and administrative areas was done. The best, lowermost org-unit of information generation was selected, after considering the availability of GIS boundaries, static nature and granularity of generated information.

7   Chapter 4. Organisation units [Internet]. Available from: https://docs.dhis2.org/2.22/en/user/html/ch04.html
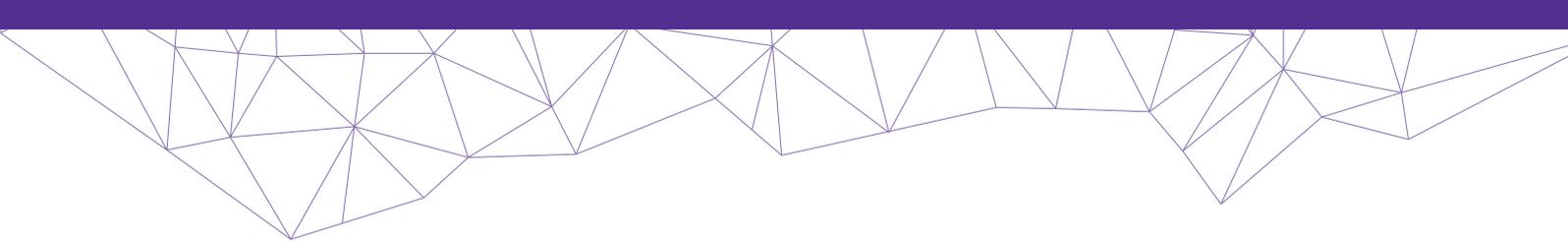
Importing VA metadata into DHIS2

Metadata refers to information about the data to be collected.[8] DHIS2 has an import/export metadata functionality that allows users to easily transfer metadata from one DHIS2 instance to another. Of relevance to event programs (such as that used for VA), exchangeable metadata includes Program, Program Stage, Data Element (tracker domain), Option Set, Option which describe the event data collect about the tracked entities.

The available metadata files on the SmartVA DHIS2 GitHub repository are as follows:[9]

- ■ Import metadata/optionset_age_category.csv

- ■ Import metadata/optionset_cause_of_death.csv

- ■ Import metadata/optionset_ICD10.csv

- ■ Import metadata/optionset_sex.csv

- ■ Import metadata/dataelements.csv

- ■ Import metadata/program.json

Configuring the script

Configuring the script according to the country-specific VA data collection methodology is a critical step. The script can be configured to automatically pull data from an ODK Aggregate server, or manually through downloaded CSV files that contain VA data under proper column headers. For automation, ODK server credentials should be inserted into the configurations, which can be found at /home/dhis/smartva-dhis2/dish.json.

```
{
  "dhis": { //DHIS2 Server Details
    "baseurl": "http://localhost:8080",
    "username": "admin",
    "password": "district"
  },
  "odk": {  //ODK Server Details
    "baseurl": "https://odk-aggregate-server.org",
    "username": "user",
    "password": "pass"
  }
}
```

As an alternative the script can manually capture VA data by addressing the downloaded ODK Aggregate CSV file as below example command., the Mortality Technical Working Group of each hospital reviews the certificates for quality and tries to mitigate any gaps in information.

```
Pipenv run python -m smartvadhis2 –manual '/media/sf_Windows_Linux_
Shared/SmartVA_SriLanka_results.csv'
```

The configuration settings of config.ini should be accessed by doing as follows.

```
sudo nano /home/dhis/smartva-dhis2/config.ini
```

8   Definition of METADATA [Internet]. [cited 2018 Nov 1]. Available from: https://www.merriam-webster.com/dictionary/metadata

9   GitHub - D4H-VA/smartva-dhis2: Verbal Autopsy data integration into DHIS2 [Internet]. Available from: https://github.com/D4H-VA/smartva-dhis2

And the following variables should be incorporated to match the country-specific VA metadata.

```
# smartva-dhis2 configuration file
# see: https://smartva-dhis2.readthedocs.io/en/latest/configuration.
html#application

[auth]
auth_file = dish.json

[logging]
logfile = smartva_dhis2.log
level = INFO

[database]
db_queries_log = false
db_name = smartva-dhis2.db

[odk]
form_id = SmartVA_SriLanka
sid_regex = ^VA_[0-9]{17}$
#sid_regex = .*

[smartva]
ignore_columns =
geography1,geography2,geography3,geography4,geography6,cause34
algorithm_version = Tariff 2.0
country = LKA
hiv = true
malaria = true
hce = true

[dhis]
program = HPrJOsYuM1K
program_stage = pQ8gaWKD3pi
api_version = 29
```

Localising the ODK Collect form

It is essential that all location data are recorded as 11-digit unique identification (uID) numbers to be transmitted to the relevant org-unit in DHIS2. Many countries participating in D4H that have implemented SmartVA initially collected location data as free-text, including Sri Lanka. Once such data have been uploaded into the ODK Aggregate server there is no mechanism for changing the location data, for example, to fix errors or standardise terms, apart from resubmitting the data. A minor typo by the VA interviewer could result in mismatching of org-units and rejection by the integration script.

Therefore, it is crucial to setup the DHIS2 org-unit structure as the first line of activity in implementing integration processes. A working solution is to provide the appearance of the relevant org-unit from the tablet/front end to the end-user (VA interviewer) and to record the relevant uID in the back end at the ODK Aggregate server. This enables automatic flow of data through ODK Briefcase and Tariff 2.0 analysis via the integration script and pushing individual cause of death data into the DHIS2 platform.
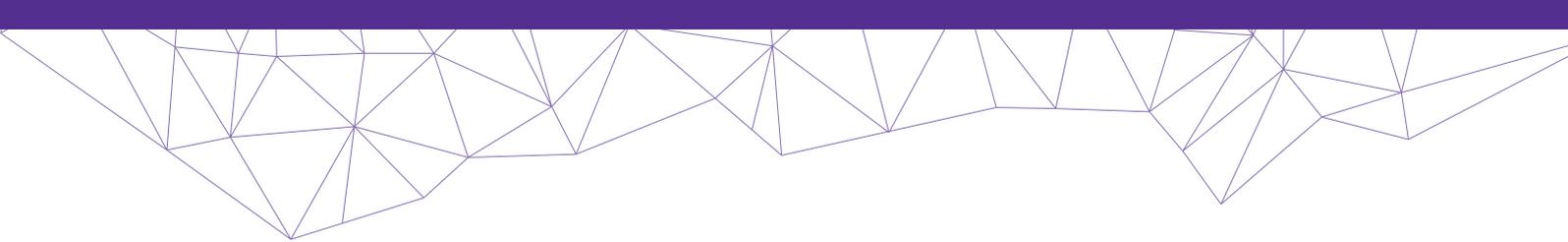
In order to achieve ODK Collect form modification to capture org-units by DHIS2, the following pre-requisites are required:

1. Organisation structure hierarchy with co-relating uID
2. Country specific ODK Collect form along with pre-compiled Excel file (i.e. the Excel file has been updated to match country specifics, like language translations) and other media files
3. ODK Aggregate server with admin rights for testing
4. Android Tablet with ODK Collect installed and configured server settings according to above ODK Aggregate server.

The selection of org-units can be done in two ways:

1. Embedding all org-unit data into the ODK Collect form itself
2. Structuring org-unit data in an external CSV media file so they can be pre-loaded upon filling a new form.

Sri Lanka has nine provinces, 25 Districts, 331 Divisional Secretariat (DS) Divisions and 14,022 Grama Niladhari (GN) divisions (equivalent to a village unit). In terms of Health boundaries after district's boundaries, there are 341 Medical Officer of Health (MOH) Areas. Public health inspector (PHI) and PHM (public health midwife) areas overlap with periodical administrative changes in the public health sector (**Figure 1** illustrates health and administrative boundaries). This introduces many challenges when attempting to embed org-units including increasing the size of the ODK Collect form main file that can slow down and complicate downloading of ODK Collect forms by end-users, especially when there is lack of proper internet connectivity. Implementing changes into the org-unit through main form modification may cause disruption in the entire data collection process by end-users as all tablets may have to be re-configured, and pre-collected data may be subjected to rejection by the ODK Aggregate server if there is mismatching data between Aggregate and the tablet. Therefore, naming org-units from an external CSV file was considered as more appropriate and practical.

Figure 1. Health and administrative boundaries in Sri Lanka
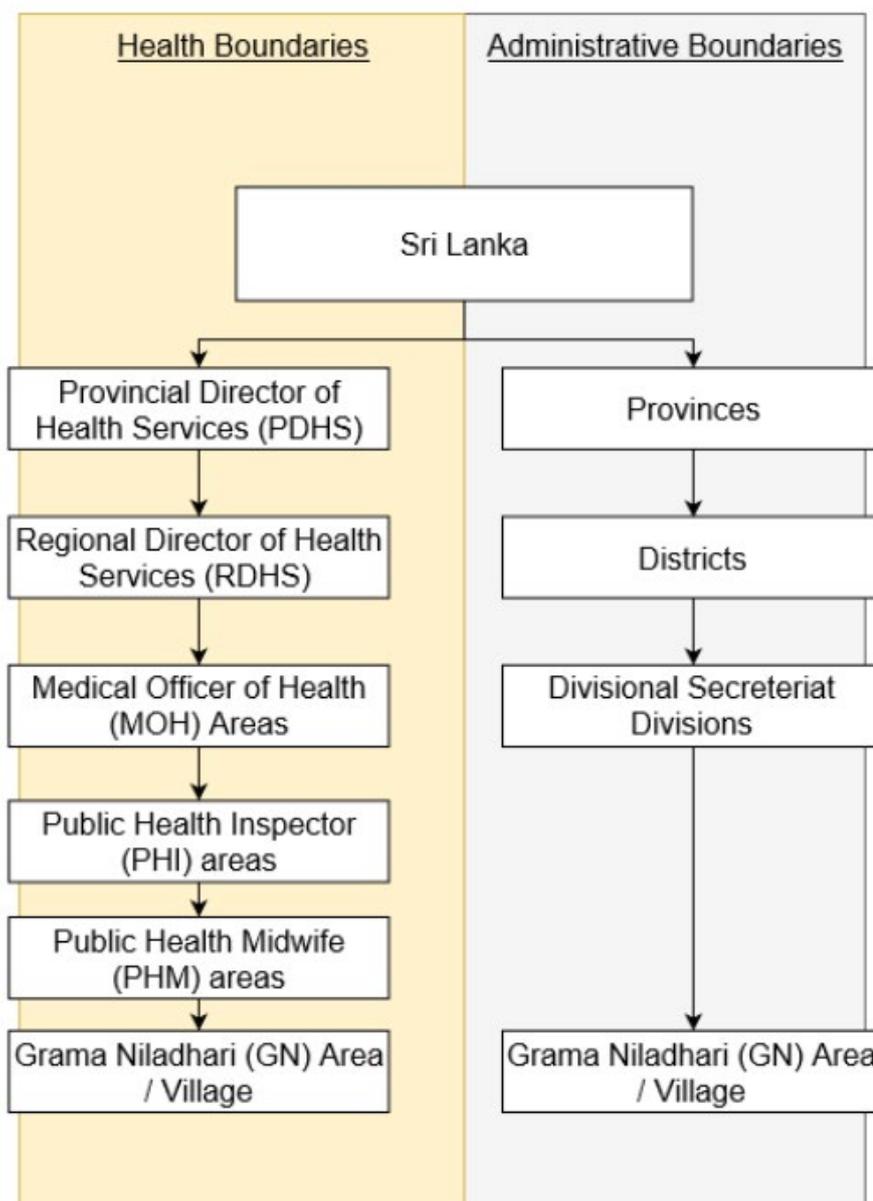
Before applying the location capture component into the main ODK Collect from of Sri Lanka, a thorough analysis of the existing structure of the ODK Collect form was done. The primary goal was to apply a minimum amount of modification into the existing form to prevent collapse of the entire data flow setup so far. As Sri Lanka is in the early stages of implementation of SmartVA, this provides a good environment to test this tool and make the changes from the beginning.

Analysis of existing ODK Collect format for Sri Lanka VA data collection

Currently, on focusing of the structure of the Sri Lanka SmartVA form location module, it has called upon locations of the pilot areas only as following structure:

1. District
2. MOH
3. PHM areas

In previous trials and discussions, it was considered using GN areas possess a definitive advantage over PHM areas in collecting the root location. A better understanding of the methodology of existing ODK Collect location capture was thoroughly studied using information from the xlsform documentation.[10]

Modification of existing ODK Collect form to collect DHIS2 uID

Based on the results of the primary analysis of the existing ODK Collect format for Sri Lanka, the following changes were incorporated. Change of location column headers to *gen_2_3#* format as follows. Since *gen_2_3#* column headers were already used by VA study ID formulation group. That group of column headers were replaced by *gen_2_2#*. Appearance arguments and constraints were added as following (**Table 2**).

Table 2 ODK form modifications for Sri Lanka

general2	Background Section
phm_name	PHM's Name?
phm_contact	PHM's contact/phone number?
gen_2_3c	District?
gen_2_3d	MOH Area?
gen_2_3e	GN Area?
gen_2_4	Address of household

10   8.3. DHIS2 Server setup [Internet]. [cited 2018 Oct 31]. Available from: https://docs.dhis2.org/2.28/en/implementer/html/install_server_setup.html

Results

Setting up the integration script

The process of setting up DHIS2 was time-consuming and required thorough troubleshooting, involving a series of trial-and-error tests to ensure that no errors occurred. After troubleshooting issues were resolved, the process of setting up DHIS2 took about 30 minutes, and was categorised into the six following phases:

1. Updating and configuring Linux OS
2. Installing Java Runtime Environment
3. Installing PostgreSQL
4. Installing and configuring Tomcat / Servlet
5. Downloading and configuring DHIS2
6. Running DHIS2.

In reference to the test cases described in the Methods section, tests 1 and 2 performed well with minimal complications, whilst tests 3 and 4 failed. Test 3, however, appeared to have a better selection of apps, improved dashboard features, and a user interface (UI). Therefore, it was recommended to proceed with DHIS2.29 in an Ubuntu 16LTS installation.

It was essential that configurations in the script be changed to map out the data transfer from CSV files and ODK server. To do that, "config.ini" (located in /home/smartva-dhis2 folder in gui) was modified in the SmartVA DHIS2 GitHub clone. Then the lowest organisational unit in the hierarchy was assigned to the column "general5: location: gen_2_3e". Next, in the config.ini under [smartva] tag ignore _columns = all tagged with geography [1-5] was added – except for the geography 5 (which is assigned for low-most data generating org-unit). Finally, ignore-columns = geography1, geography2, geography3, geography4, cause34. Change of country = LKA was also done.

A slight mismatching of the documentation and metadata available in the SmartVA-DHIS2 GitHub was observed,[11] where the documentation noted seven metadata files to be imported whereas the repository only had six. The dashboard metadata file did not appear to be available. Upon trial, however, all test cases except for tests 3 and 4 were able to import metadata files successfully. It was learned that best practice involves doing a 'dry run' before importing metadata, and that relevant organisational units, program metadata, and data elements will not appear just after metadata are imported. Running analytics and maintenance during this phase are thus necessary.

Test 3 – in which DHIS2.30 was installed – exhibited problems with importing metadata, particularly with importing organisational unit XMLs and data element CSV files which were successfully imported in previous tests. The metadata import application in DHIS2.30 appeared to have had undergone a total makeover without these changes having been documented in official releases of DHIS2 documentation.

Upon analysis of survey declaration of location capture variables in the existing ODK Collect form, it was observed in the "appearance" column of the survey form that minimal search syntax had been used in order for the ODK Collect form in the tablet to appear like a drop-down menu. Upon final testing with the ODK Aggregate set-up, it was noted that the integration script does not cooperate with any name tags (other than "gen_2_2d" to

11   GitHub D4H-VA/smartva-dhis2: Verbal Autopsy data integration into DHIS [Internet]. Available from: https://github.com/D4H-VA/smartva-dhis2
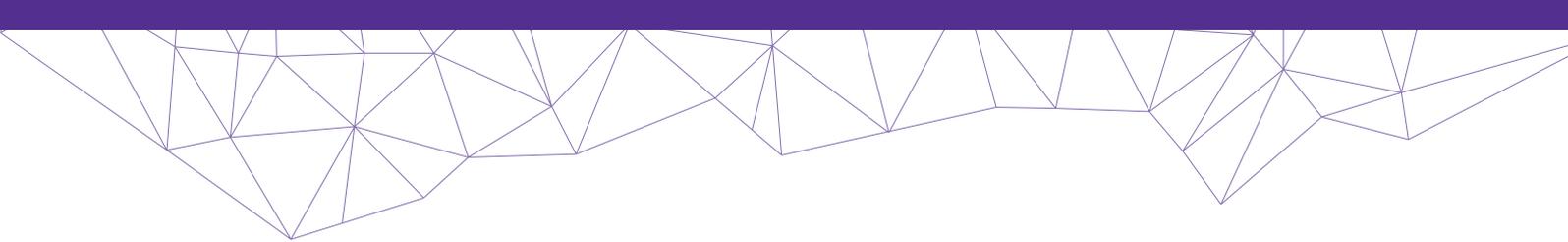
"gen_2_2f") when it comes to location selection. Because GN areas were assigned – instead of PHM areas – were assigned as the lower-most hierarchical unit, these areas were called upon from an external CSV file where all Sri Lankan location data can be captured in a way where the front end displays the organisational unit name, and the back end captures the uID along a cascading selection methodology as described in XLS forms official documentation.[12]

Running the script

Running Linux commands had to be done carefully. Most of the commands were case-sensitive, and even an extra space in the code could change the outcome. After several trials, many errors were neutralised using typos in Linux commands. The following errors appeared to be common:

Command Line Interface (CLI) errors

It is advised to do an apt-get update before running the script in a new server environment. A common error was the failing of "pipenv install -ignore-pipfile –deploy", resulting in "pipenv:command not found", which can be remediated by invoking the following command and re-trying the process:

```
sudo pip3 install pipenv
```

Regex error

When pushing data to DHIS2, the following error was given for all 469 data entries:

```
ERROR 180808 16:48:59  Error (609): [sid] does not match regex expression
[run:96]
```

After troubleshooting and analysis of the entire dataset, it was found that the configuration in /home/dhis/smartva-dhis2/config/ini was the root cause of this error.

```
[odk]
form_id = SmartVA_SriLanka
sid_regex = ^VA_[0-9]{17}$
#sid_regex = .*
```

It is important to either enter a proper regex expression or uncomment "sid_regex = .*" to allow any regex and comment(#) the defined expression to inactivate the relevant code on compiling.

Invalid org-unit error

During the initial testing phase, the following error was given for all 469 entries

```
ERROR  180808 17:23:28  Error (612): orgunit UID is not a valid UID
[run:96]
```

To rectify this, the column headers were changed in the Sri Lanka SmartVA results CSV file. However, the error continued to appear until the location column headers were changed to a different format, as described earlier in the Methods section (Table 2).

---

12   XLSForm.org [Internet]. Available from: http://xlsform.org/en/

Duplicate VA study ID

Duplicate VA study IDs appeared on the existing ODK Collect form used for the pilot phase in Sri Lanka. It was determined that these duplicate IDs were generated through incorporation of variables like 'year', numbering mechanisms related to 'province', 'district', 'MOH area', 'PHM area', and 'death certificate number'. For 'death certificate number', it was indicated to use '999' if unknown. This has resulted in many duplicate entries during VA data collection and has complicated the integration module by rejecting certain VA entries by the script, as the script states that the entire VA entry is a duplicate.

VA dashboards

<div style="float:left; font-style:italic;">Customised DHIS2 dashboards can be created.</div>

After all VA data has been pushed onto DHIS2, customised DHIS2 dashboards can be created according to end-user preferences by using the instructions provided by the official user guide of DHIS2 on dashboards.[13] Dashboards created in the past include 'COD distribution of Sri Lanka for 2018', and 'VA distribution by sex in Sri Lanka for 2017-18', 'comparison of VA data collected 2017-18 and its geographical distribution at provincial level and GN level'. The dataset used comprised of 469 VA records collected during the pilot phase of SmartVA implementation in Sri Lanka.

These dashboards can be easily customised, shared, embedded on websites, and updated in real-time. It is also necessary to note that such custom dashboards can be prepared using the same dataset according to area-specific managers using the 'user access control' capabilities of DHIS2. **Figures 2 & 3** show sample dashboards generated from the test dataset.

Figure 2 SmartVA-DHIS2 dashboard showing COD distribution and VA distribution by sex
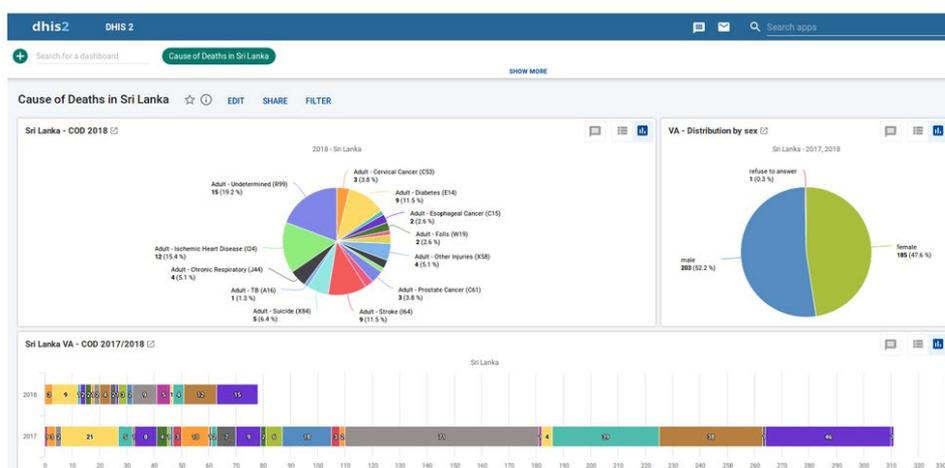
13   DHIS2 Installation Guide for Ubuntu 16.04 LTS [Internet]. Available from: http://dhis2.blogspot.com/2018/02/dhis2-installation-guide-for-ubuntu.html

Figure 3 SmartVA-DHIS2 dashboard showing geographical distribution of VAs
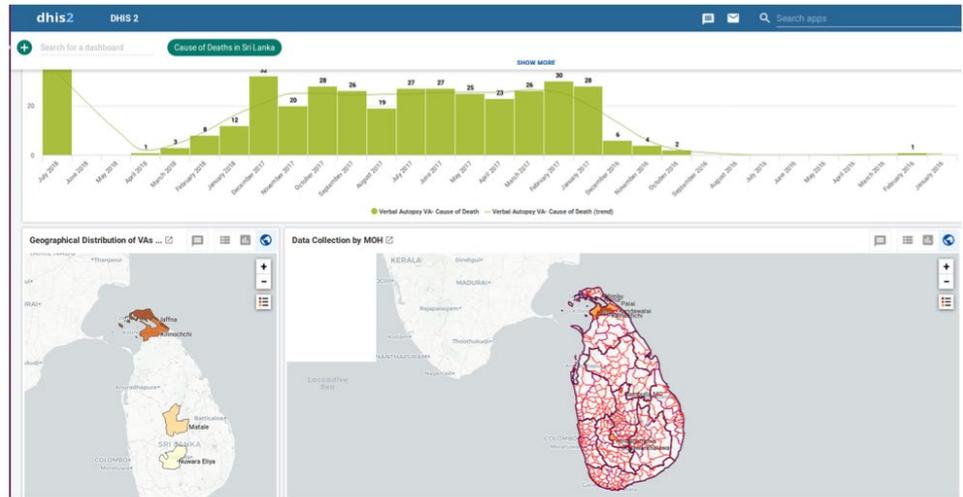
Summary

The exponential growth of DHIS2 implementation over HISP networks across the globe means that many countries are now interested in adopting DHIS2 as their primary public healthcare platform. The SmartVA-DHIS2 integration script is a step towards populating dashboards across SmartVA implementations, which allows end-users, managers, researchers, and administrators to visualise, analyse, interpret, and disseminate data whilst monitoring data collection in real-time. All of these functions are complemented by DHIS2's facilities ensuring data security.

Automating a data flow that was already in operation was difficult.

Automating a data flow that was already in operation is not a straightforward task. The main challenges include the lack of a real-time testing server environment, the need for thorough (and often time-consuming) trial-and-error testing to ensure accuracy of the automated outputs, and the need for customisation of the integration module into a country-specific environment – given that countries all use different software setups, network protocols, and server environments.

Recommendations

Having an in-depth analysis and understanding of existing workflows is critical.

We have several basic recommendations for addressing the challenges:

- Perform a thorough in-depth analysis of existing workflow involving movement of every data element
  - This analysis should include basic concepts of information science such as what type of data, where and when the data are generated, and where the data go and when they should be delivered.
- Identify the possible links of automation to generate an automation workflow
- Following the analytical phase, research the existence of any similar use-cases and patterns of automation.

CRVS Fellowship reports and profiles

Related resources and products

University of Melbourne, D4H Initiative, CRVS Knowledge Gateway: Library

https://crvsgateway.info/library

Action guide on automated VA training and rollout. CRVS action guide.

Challenges associated with automated VA training and rollout. CRVS development series.

CRVS country overview: Sri Lanka. CRVS summaries.

Developing a mobile app for doctors to improve the recording of cause of death in Sri Lanka. CRVS summaries.

Improving the quality of mortality statistics in Sri Lanka. CRVS summaries.

Intervention: Automated verbal autopsy. CRVS summaries.

SmartVA DHIS2 integration. Video.

SmartVA: Interviewer's manual. CRVS resources and tools.

SmartVA: Technical user guide (V1.0). CRVS resources and tools.

Sri Lanka: Strengthening the quality and availability of mortality statistics. CRVS development series.

University of Melbourne, D4H Initiative, CRVS Knowledge Gateway: Learning Centre

https://crvsgateway.info/learningcentre

Topic 3: CRVS processes.

Topic 4: Cause of death in CRVS.

University of Melbourne, D4H Initiative, CRVS Knowledge Gateway: Courses

https://crvsgateway.info/courses

SmartVA.

Further reading

de Savigny D, Riley I, Chandramohan D, Odhiambo F, et al. Integrating community based verbal autopsy into civil registration and vital statistics (CRVS): system-level considerations. *Global Health Action* 2017; 10(1): 1272882.

Flaxman A, Stewart A, Joseph JC, Alam N, et al. Collecting verbal autopsies: improving and streamlining data collection processes using electronic tablets. *Population Health Metrics* 2018; 16:3.

The program partners on this initiative include: The University of Melbourne, Australia; CDC Foundation, USA; Vital Strategies, USA; Johns Hopkins Bloomberg School of Public Health, USA; World Health Organization, Switzerland.

Civil Registration and Vital Statistics partners:

For more information contact:

CRVS-info@unimelb.edu.au
crvsgateway.info

unimelb.edu.au