SIDAInfo-DHIS2 Interoperability

**Administrator’s Manual**

English version

June 2022

# **Contents**

**Section 1: Introduction** Page 3

A description of the two systems connected by the SIDAInfo-DHIS2 interoperability, as well as a description of the architecture of the solution.

**Section 2: Installing the interoperability scripts** Page 6

Instructions for installing the interoperability script on a computer for the first time.

**Section 3: Routinely updating the SIDAInfo data held in DHIS2** Page 7

Instructions on how to maintain this interoperability on a regular basis (ie how to routinely update the copy of SIDAInfo data that is held in DHIS2).

# **Section 1: Introduction**

**1.1 BACKGROUND**

**SIDAInfo** is an electronic medical record that is used by all **ART / PMTCT** sites in Burundi, including the 8 sites supported by PSI Burundi. Individual records are maintained for all patients as they visit, and all consultations, prescriptions, diagnostic tests and other key events are captured in the system.

**PSI’s DHIS2 instance** (https://data.psi-mis.org) is a global server that is used by PSI country programs all over the world, including PSI Burundi, to manage data and reporting. PSI’s DHIS2 instance already has programs to capture **prevention, testing and counselling** data, but PSI’s DHIS2 instance does not directly capture ART or PMTCT data.

The **SIDAInfo-DHIS2 interoperability** script connects these two systems with a one-way data flow that loads a subset of essential patient-level data from SIDAInfo into three Tracker programs in DHIS2:

* A PMTCT Mother Tracker program
* A PMTCT Child Tracker program
* An ART Tracker program

These Tracker programs are **read-only** – in other words, data is never modified in DHIS2 itself, it is used solely for analytical purposes. (All updates and correction are made to the source records in SIDAInfo, and then re-uploaded.)

Based on these DHIS2 Tracker programs, a range of DHIS2 indicators have been developed, which automatically calculate key **standard ART and PMTCT indicators**, enabling the development of dashboards and other analytics in DHIS2. (A full list of these indicators can be found in the Specification document.)

A wide range of **data validation routines** are also included in the interoperability script, which carefully check the SIDAInfo patient data as it is loaded. (A full list of these data validation checks can be found in the **Specification** document.) If an error is detected, there are two possible outcomes:

1. For a small number of serious data quality issues which affect the integrity of patient data, the affected patients are excluded from the upload to DHIS2 (a log is generated listing these patients, for use by system administrators).
2. For the majority of more minor validation errors, the patient data is still uploaded to DHIS2, but an error message is displayed in DHIS2 highlighting the errors, so that site staff can review, correct and re-upload the patient data.

**1.2 ARCHITECTURE**

As noted above, the SIDAInfo-DHIS2 interoperability involves a **one-way flow of data** from the source systems (SIDAInfo installations at each site) into PSI’s DHIS2 instance. The diagram below shows the flow of data from SIDAInfo into DHIS2:

**PSI’s DHIS2 server**(global, holds all sites)

**SIDAInfo installation**  
(local, a single site)

CSV files exported monthly or weekly by site staff and emailed to PSI Burundi team.

**CSV file**

(all patient data held by one site)

**Interoperability  
scripts**

Step 1:   
**sidainfo-dhis2.exe retrieve**Interoperability script pulls down all existing   
SIDAInfo data already   
on the DHIS2 server   
to enable a   
comparison.

Step 2:   
**sidainfo-dhis2.exe generate**  
Latest SIDAInfo   
data is linked to   
DHIS2 patient   
UIDs and prepared   
for analysis.

Automated data   
validation routines   
are also run, and   
a list of data quality   
errors is generated.

Step 4:  **sidainfo-dhis2.exe upload**  
Data in DHIS2 is automatically updated (updating only records that have changed).

Step 3:  **sidainfo-dhis2.exe diff**  
Existing and new data are compared and a small list of changes is identified.

A few important notes on this architecture:

* The interoperability process is **incremental**: in order to make the updates as fast and efficient as possible, the existing data on DHIS2 and the new data received from the site is compared, and **only changes are uploaded to DHIS2**. So if a patient has additional activity or a new patient has been added, this will be uploaded to DHIS2; however, if the patient’s record has remained exactly the same since the last time the interoperability scripts were run, then nothing will be uploaded to DHIS2. This is *significantly* faster than re-uploading all SIDAInfo data every time, as data can be very rapidly read from the DHIS2 API, but updates to the DHIS2 API are relatively slow.
* Because this incremental process involves comparing SIDAInfo and DHIS2 data every time the interoperability scripts run, it is a **very** **robust interoperability process** which protects the integrity of SIDAInfo data held in DHIS2. If, for example, DHIS2 is rolled back to an earlier backup, losing SIDAInfo data that had already been uploaded, then the incremental process will detect this and re-upload the missing data. Likewise, if data is deleted from a SIDAInfo site after it has been uploaded to DHIS2, the incremental process will detect this and then also remove that data from DHIS2. (Note that all deletes in DHIS2 are ‘soft deletes’, which means a hidden record of all deleted data is stored in DHIS2.)
* The interoperability scripts currently target PSI’s global DHIS2 instance, but because this has been designed as a generic interoperability architecture, it could in fact be used (with appropriate modifications) to share SIDAInfo data with *any* DHIS2 instance, potentially including the Ministry of Health’s own DHIS2 servers.

# **Section 2: Installing the interoperability scripts**

The SIDAInfo-DHIS2 interoperability consists of a folder (containing a script plus two files) that needs to be copied onto a computer with an internet connection and at least 100Gb of hard disk space available. Once this folder has been installed, the script can then be used to load SIDAInfo CSV files and upload them to DHIS2. The script that is provided with this documentation is designed to be run on a **Windows computer**. (But if needed, the source code could be recompiled to also run on a Mac or Linux computer.)

If you are using a new computer for the first time to run the interoperability script, you will need to first copy the folder containing the script onto that computer, and then ensure that all configurations are correct. The steps for doing this are outlined below:

1. **Unzip the sidainfo-dhis2.zip file onto your Windows computer:** find an appropriate location on your hard drive (it can be anywhere on your hard drive that has read-write access), copy over the installation file called sidainfo-dhis2.zip, and extract/unzip it.
2. **Check that all necessary files and folders are present:** the following items should be showing in the new sidainfo-dhis2 folder:

* sidainfo-dhis2.exe
* ou\_mapping.json
* config.json

Check the ‘properties’ of the folder, to ensure users have access and it isn’t read-only.

1. **Set up the server connection in the config.json file:** update the config.json file with the credentials to access the dhis2 instance you are targeting. By default, the config file is setup to use PSI’s “Dev” server (dev.psi-mis.org) for testing purposes; before you go live with this interoperability, you will need to replace it with PSI’s “Data” (live/production) server (https://data.psi-mis.org). You will also need to add a username and password for a user that has full access (read-write) to the SIDAInfo Tracker programs.
2. **Configure the organization unit mappings in the ou\_mapping.json file:** in order to ensure the SIDAInfo data is linked to the correct site in DHIS2, each site code in SIDAInfo needs to be mapped to a DHIS2 organization unit UID. The ou\_mapping.json file contains a list of these mappings – you can add as many lines as necessary. (It is very important that *all* PSI-supported sites are included in the mapping.) The format is "SIDAInfo site code": "DIS2 org unit UID", eg:  
    "17020203": "DLsHsaJhtnk"  
   **Important:** there must be a comma after each line of the mapping except the last line.

Your SIDAInfo-DHIS2 interoperability scripts should now be ready to run on a routine monthly/weekly basis – see the next section for instructions on how to do this.

# **Section 3: Routinely updating the SIDAInfo data held in DHIS2**

This section describes how to routinely run the interoperability process in order to update DHIS2 with the latest SIDAInfo patient data from each site. It assumes that you have already installed the interoperability scripts on a Windows PC (see section 2).

In line with the architecture shown in section 1, maintaining this interoperability involves four main steps:

1. Retrieve existing SIDAInfo data from the DHIS2 server for comparison purposes.
2. Prepare the latest SIDAInfo CSV files sent by sites.
3. Run the automated comparison of the latest SIDAInfo data and existing DHIS2 data, generating a file that contains the changes that need to be uploaded to DHIS2.
4. Run the upload to incorporate these changes into DHIS2.

**Important: all of the above steps must be run on the same date for each site.** Do not start this process for a site on one day and try finishing it on the next day (and also do not work past midnight!) – the process must all be completed on the same day or else start again from scratch on the next day. (It is, however, fine to upload some sites one day, and other sites the next day.)

Your **antivirus may warn you** against running the commands below – if it does, please ignore the warnings / click ‘yes’ to continue running it.

**Step 1: Retrieve existing SIDAInfo data from the DHIS2 server for comparison purposes**

1. Double check that the config.json file contains the credentials to access the dhis2 instance, and that the ou\_mapping.json file contains a mapping for the site you are uploading. If not, see section 2 of this manual for instructions on how to update them.
2. Check the latest set of CSV files that you have received from this site and **note down the date shown in the filename**. This is the “CSV file date”, which you will use during various steps below. You will need to **use this date in reverse order (YYYY-MM-DD) for the sidainfo-dhis2 script**, eg:

31-01-2022 becomes 2022-01-31

1. Open a Command Prompt in windows and use the “cd” command to change to the sidainfo-dhis2 folder you have set up. Your command prompt will differ depending on where you have installed the file, but it should look something like this:  
    C:\Users\...\**sidainfo-dhis2>**
2. Run the program sidainfo-dhis2.exe, with the command retrieve and using the arguments  
    --site=[site code] (eg 17020203)  
    --export\_dump\_date=[CSV file date] (eg 2022-01-31).  
   For example, the command might look like:  
    sidainfo-dhis2.exe retrieve --site=17020203  
    --export\_dump\_date=2022-01-31  
   This will download existing SIDAInfo data already in DHIS2 (for comparison purposes), putting the following files into the folder PREVIOUS\_DHIS2\_data\[site code]\_[CSV file date]:

* enfant.json
* mere.json
* tarv.json
* teis.json
* previous\_all\_patient\_index.json

Also, log files will be saved in the log folder

* retrieve.log
* retrieve-error.log

(To make sure no issues have been encountered with this process, open the retrieve-error.log file and see if any error messages are shown.)

**Step 2: Prepare the latest SIDAInfo CSV files sent by sites**

1. **Gather the latest site CSV files and check them:** each PSI-supported site will use a function in SIDAInfo web to export a set of CSV files containing the latest patient data and will send this to the PSI team ready for upload to DHIS2. The site may send 7 separate .txt files, but they may also come as a single .zip or .rar file (in which case, extract the .zip or .rar to produce the 7 separate .txt files). Check that all 7 files are present:

* ftp\_[sitecode]\_[date]\_ADMISSION\_DETAIL
* ftp\_[sitecode]\_[date]\_CONSULTATION
* ftp\_[sitecode]\_[date]\_DEPISTAGE\_DETAIL
* ftp\_[sitecode]\_[date]\_ENFANT\_PTME
* ftp\_[sitecode]\_[date]\_FEMME\_ENCEINTE
* ftp\_[sitecode]\_[date]\_FileActive
* ftp\_[sitecode]\_[date]\_ADMISSION\_DETAIL

If any files are missing, do not proceed – contact the site and ask them to re-send the files.

1. **Copy the 7 new CSV files into the new sub-folder that has been created:**  
   within the sub-folder \sidainfo-dhis2\NEW\_SIDAINFO\_data, a new empty folder will have been automatically created with the site code and the “CSV file date”, for example:  
    sidainfo-dhis2\NEW\_SIDAINFO\_data\17020203\_2022-01-31\

Copy the 7 new CSV (.txt) files into this new sub-folder.

1. **Process the 7 new CSV files ready for analysis:** taking note again of the “CSV file date”, run the program sidainfo-dhis2.exe, with the command generate and using the arguments--site=[sitecode] (like 17020203)  
    --export\_dump\_date=[CSV file date](like 2022-01-31)

For example, the command might look like:

sidainfo-dhis2.exe generate --site=17020203   
--export\_dump\_date=2022-01-31

Once you have run this command, the following outputs will be generated:

* A set of new .json files will be automatically placed in the sub-folder GENERATED\_data\[sitecode]\_[CSV file date]\teis ready for comparison with the existing data that has been retrieved from DHIS2.
* A file called generated\_all\_patient\_index.json will be automatically placed in the same sub-folder, with an index of the patient data.
* A set of logs will be added to the log folder containing a list of data validation errors in English (generation-error.log) and French (generation-error-fr.log).

**Step 3: Run the automated comparison of the latest SIDAInfo data and existing DHIS2 data**

Now that both existing data already in the DHIS2 server and the new data to be uploaded are ready, we can compare them to identify the list of changes to DHIS2 that are needed.

1. Run the program sidainfo-dhis2.exe, with the command diff and using the arguments  
    --site=[sitecode] (like 17020203)  
    --export\_dump\_date=[CSV file date](like 2022-01-31)

For example, the command might look like:  
 sidainfo-dhis2.exe diff --site=17020203  
 --export\_dump\_date=2022-01-31

The following outputs will be generated:

* A file that contains a list of the actions required to update DHIS2 with the latest version of SIDAinfo data. The file is generated at GENERATED\_data\[sitecode]\_[CSV file date]\actions.json
* A log file: log\diff.log

**Step 4: Run the upload of these changes into DHIS2**

1. Run the program sidainfo-dhis2.exe, with the command upload and using the arguments  
    --site=[sitecode] (like 17020203)  
    --export\_dump\_date=[CSV file date](like 2022-01-31)

For example, the command might look like:  
 sidainfo-dhis2.exe upload --site=17020203  
 --export\_dump\_date=2022-01-31

It will take a little while for the list of actions generated in step 3 to run, updating the DHIS2 server with all necessary changes. The following outputs will be generated:

* A log file: log\upload.log

**The interoperability process is now complete!**

If you want to double-check that all updates have been correctly made on DHIS2, you can repeat step 1 (Retrieve existing SIDAInfo data from the DHIS2 server) and step 3 (Run the automated comparison) described above. This will re-do the comparison, adding entries to the diff.log file in the ‘log’ folder and regenerating the ‘actions.json’ file. If the update is successful, then the ‘actions.json’ file should be empty, as there are no new updates to be made (data on the DHIS2 server is up-to-date).