



**SOP: Data Cleaning using DWAPI Application**

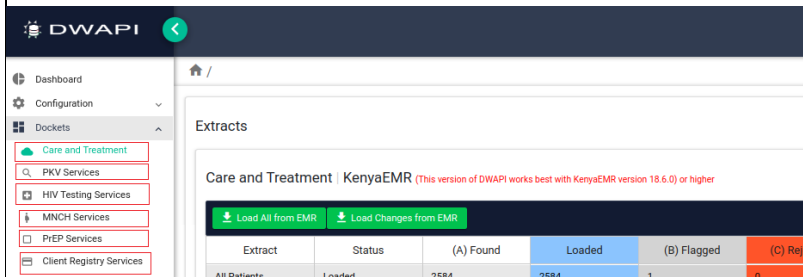
**Applies to: DWAPI 3.1.0.7 and above**

**Revised: July 2023**

<b>Tasks:</b>	Conduct EMR data cleaning using DWAPI data validation tool.
<b>Objective:</b>	How to use DWAPI to clean EMR data at the facility level.
<b>Who:</b>	Clinicians/Data Staff/M&E
<b>Required Materials:</b>	Functional DWAPI tool, Functional EMR, Source clinical documents/ Registers, Internet Connectivity
<p><b>Summary:</b></p> <p><i>Besides the primary function of EMR data upload into the National data warehouse, DWAPI tool also offers crucial data validation and cleaning feature, which if properly and constantly utilized, may lead to data quality improvements in a facility. This document provides a stepwise guide on how to use DWAPI data validation feature to achieve improved data quality in EMR implementing facilities. This procedure can be performed before or after submitting data into the Data Warehouse (DWH).</i></p> <p><b>NOTE: Data can be uploaded into Datawarehouse as many times are necessary depending on data cleaning iterations undertaken at the facility.</b></p>	
Action	Screen shot

### 1. Navigate to DWAPI Dockets:

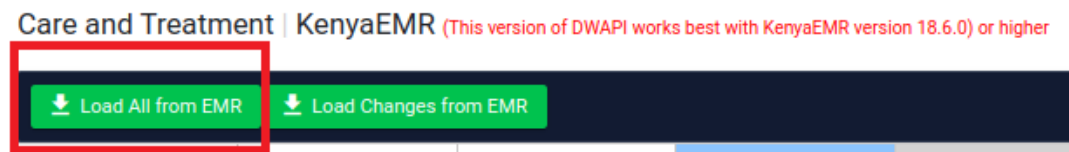
Navigate and select from any of the available docket of interest in DWAPI application. E.g. Care & Treatment, PKV Services, HIV Testing Services, MNCH Services, PREP Services, Client Registry Services as highlighted. For illustration purposes in this document, we are going to use **Care and Treatment** Docket. For details on how to Load and send data to the warehouse using the DWAPI tool.



### 2. Load data (extraction):

Load the EMR data into the selected docket by clicking on the **“Load From EMR”** button. Wait until all the datasets are populated and the status for all the datasets is shown as **“Loaded”** as shown.

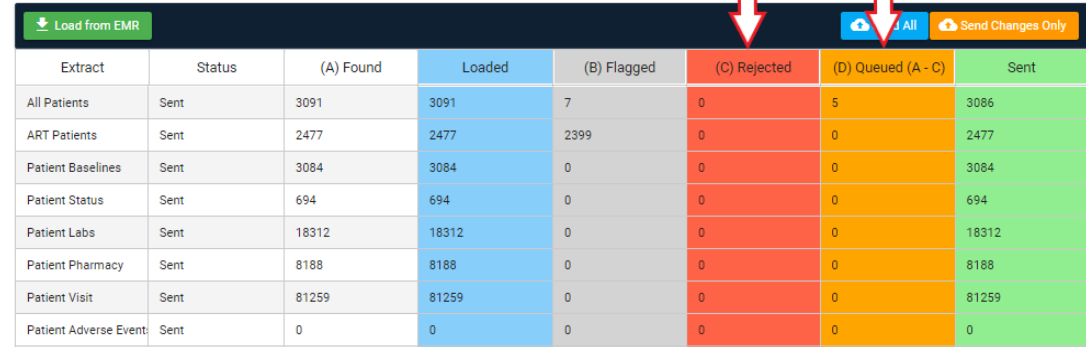
#### Extracts



### 3. View Flagged/ Rejected records:

Once the data is successfully loaded into each dataset, check under “**Flagged**<sup>1</sup>” and “**Rejected**<sup>2</sup>” columns for the number of records flagged or rejected by the application.

Care and Treatment | KenyaEMR (This version of DWAPI works best with KenyaEMR version 17.3.5) or higher



Extract	Status	(A) Found	Loaded	(B) Flagged	(C) Rejected	(D) Queued (A - C)	Sent
All Patients	Sent	3091	3091	7	0	5	3086
ART Patients	Sent	2477	2477	2399	0	0	2477
Patient Baselines	Sent	3084	3084	0	0	0	3084
Patient Status	Sent	694	694	0	0	0	694
Patient Labs	Sent	18312	18312	0	0	0	18312
Patient Pharmacy	Sent	8188	8188	0	0	0	8188
Patient Visit	Sent	81259	81259	0	0	0	81259
Patient Adverse Event	Sent	0	0	0	0	0	0

**Note:** Scroll down under ‘**KEY**’ to see the definition of each color-coded status.



KEY

**\*Found** => Number of records available in EMR for each dataset.

**\*Loaded** => Number of valid records loaded from EMR for each dataset

**\*Rejected** => Records in the extract that will not be sent to the warehouse because they belong to a patient(s) with errors.

**\*Flagged** => Number of Unique patient records with one or more issues that need to be cleaned in the source EMR. These issues are recorded for review in the validation summary section. For better data quality, site data managers/HRIOs should clean these records before sending to the datawarehouse

**\*Queued** => Number of records in each dataset ready for sending to the data warehouse. These records will include flagged records that have not been cleaned.

**\*Sent** => Total number of records sent to the data warehouse.

<sup>1</sup> **Flagged record:** These are records with notable inconsistencies or which have some variables missing, but may still be allowed to pass (uploaded into DWH)

<sup>2</sup> **Rejected record:** These are set of records that lack some very critical information i.e. age, patient ID, gender. Such records are NOT permitted to be uploaded into DWH

#### 4. List of Flagged/Rejected records

After successfully ascertaining the number of records flagged or rejected, it is time to identify individual records and the reason provided. Scroll all the way down under “**Extract Details**” section.

Select the affected dataset (in this example, it is **ART Patient** - see 1). Click on **Validation Summary** – indicated 2 as shown

Extract Details

All Patients **ART Patients** Patient Baselines Patient Status Patient Labs Patient Pharmacy Patient Visit Patient Adverse Events

ART Patients Records **ART Patients Validation Summary** ART Patients Send Summary

Export

Patient ID	Summary	Extract	Field	Type	Start Regimen	Start ART Date	Start Regimen Line
100	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	TDF+3TC+EFV	2012-05-14T00:00:00	First line
1387502243	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2018-12-04T00:00:00	First line
60401000512	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	D4T+3TC+NVP	2009-12-01T00:00:00	First line
60401000234	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	AZT+3TC+NVP	2011-05-21T00:00:00	First line

#### 5. List records with issues

Under Validation summary, a list of flagged records will be shown. About eight columns are displayed, but of key interest are the **Patient ID** (1) and the **Summary** (2) as shown.

The summary column indicates the issue detected with the record while the Patient ID provides the identifier to be used for searching the affected patient in EMR.

#### Extract Details

All Patients ART Patients Patient Baselines Patient Status Patient Labs Patient Pharmacy Patient Visit Patient Adverse Events

ART Patients Records ART Patients Validation Summary ART Patients Send Summary

Export

Patient ID	Summary	Extract	Field	Type	Start Regimen	Start ART Date	Start Regimen Line
100	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	TDF+3TC+EFV	2012-05-14T00:00:00	First line
1387502243	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2018-12-04T00:00:00	First line
60401000512	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	D4T+3TC+NVP	2009-12-01T00:00:00	First line
60401000234	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	AZT+3TC+NVP	2011-05-21T00:00:00	First line

## 6. Export Data

It is recommended that you export the entire list into a spreadsheet (Excel) for easier manipulation or assignments to various individuals participating in the data cleaning exercise. To export, click on “**Export**” button. Wait for the export to download. Check in downloads folder for the file named *download*. Proceed with data cleaning based on the error reported under **Summary** column.

3

Patient ID	Summary	Extract	Field	Type	Start Regimen	Start ART Date	Start Regimen Line	DOB	Gender	Patient Source
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	TDF+3TC+EFV	2012-05-14T00:00:00	First line	1986-06-15T00:00:00	F	OPD
	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2018-12-04T00:00:00	First line	1986-06-15T00:00:00	F	OPD
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	D4T+3TC+NVP	2009-12-01T00:00:00	First line	1959-03-05T00:00:00	F	VCT
	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	AZT+3TC+NVP	2011-05-21T00:00:00	First line	1984-06-15T00:00:00	F	PMTCT
	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	3TC+NVP+TDF	2011-06-20T00:00:00	First line	2009-06-15T00:00:00	F	VCT
	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2018-11-28T00:00:00	First line	1986-03-15T00:00:00	M	OPD
	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2015-07-08T00:00:00	First line	1990-07-22T00:00:00	F	PMTCT
	LastVisit is missing	TempPatientArtExtracts	LastVisit	Logical	TDF+3TC+EFV	2019-01-08T00:00:00	First line	1996-06-15T00:00:00	F	Maternal and child health
	StartARTDate after last visit date	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2017-01-10T00:00:00	First line	1974-06-15T00:00:00	M	OPD
	Last visit date before Enrollment date	TempPatientArtExtracts	LastVisit	Logical	D4T+3TC+NVP	2009-02-24T00:00:00	First line	1976-06-15T00:00:00	F	VCT
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	D4T+3TC+NVP	2014-08-18T00:00:00	First line	1995-11-27T00:00:00	F	PMTCT
	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	AZT+3TC+EFV	2013-03-09T00:00:00	First line	1981-08-15T00:00:00	F	OPD
	LastVisit is missing	TempPatientArtExtracts	LastVisit	Logical	TDF+3TC+EFV	2011-04-19T00:00:00	First line	1982-06-22T00:00:00	M	
	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2016-09-02T00:00:00	First line	1961-06-15T00:00:00	M	OPD
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	AZT+3TC+NVP	2011-07-28T00:00:00	First line	1982-09-13T00:00:00	F	PMTCT
	Last visit date before Enrollment date	TempPatientArtExtracts	LastVisit	Logical	ABC+3TC+LPV/r	2014-06-09T00:00:00	First line	2013-05-10T00:00:00	M	PMTCT
	LastVisit is missing	TempPatientArtExtracts	LastVisit	Logical	D4T+3TC+NVP	2010-08-25T00:00:00	First line	1986-05-06T00:00:00	F	VCT
	Last visit date before Enrollment date	TempPatientArtExtracts	LastVisit	Logical	TDF+3TC+EFV	2015-04-08T00:00:00	First line	1994-06-15T00:00:00	F	PMTCT
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	AZT+3TC+NVP	2011-04-20T00:00:00	First line	1978-06-15T00:00:00	M	VCT
	StartARTDate after last visit date	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2015-03-12T00:00:00	First line	1989-03-17T00:00:00	F	OPD
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	AZT+3TC+NVP	2011-05-21T00:00:00	First line	1954-06-08T00:00:00	F	VCT
	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2015-05-07T00:00:00	First line	1989-05-03T00:00:00	F	VCT
	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	AZT+3TC+NVP	2009-03-24T00:00:00	First line	1987-08-28T00:00:00	F	OPD
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	AZT+3TC+NVP	2011-09-30T00:00:00	First line	1962-06-15T00:00:00	F	
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	ABC+3TC+LPV/r	2015-08-18T00:00:00	First line	2014-08-15T00:00:00	M	Under five clinic
	StartARTDate before patient registered	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2006-06-17T00:00:00	First line	1981-06-15T00:00:00	M	OPD
	StartARTDate after Exit Date	TempPatientArtExtracts	StartARTDate	Logical	TDF+3TC+EFV	2016-08-08T00:00:00	First line	1980-06-15T00:00:00	F	VCT
	LastVisit is missing	TempPatientArtExtracts	LastVisit	Logical	TDF+3TC+EFV	2017-05-11T00:00:00	First line	1968-06-15T00:00:00	F	OPD
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	D4T+3TC+EFV	2009-07-20T00:00:00	First line	1969-12-15T00:00:00	M	IPD - Adult
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	D4T+3TC+EFV	2013-05-24T00:00:00	First line	1991-06-15T00:00:00	M	TB
	ExpectedReturn is missing	TempPatientArtExtracts	ExpectedReturn	Logical	3TC+NVP+TDF	2011-06-20T00:00:00	First line	2009-06-15T00:00:00	F	VCT

<sup>3</sup> Patient ID column in the above spreadsheet has been greyed out intentionally.

## 7. Clean data (in EMR)

After obtaining a list of flagged /rejected records and the specific issues associated with each record as explained above, it is time to revert to EMR and do the actual data cleaning. The process of data cleaning at this stage is relatively direct.

### Steps:

- i. *Log into EMR with the right credentials*
- ii. *Search for the affected client record using the Patient ID obtained from DWAPI.*
- iii. *Based on the validation summary message, edit the section of the record as appropriate with the correct information. **(It is strongly recommended that you refer to patient physical file to obtain the correct information)***
- iv. *Save the record after editing.*

### NB:

Repeat the four steps until all the affected records are accurately corrected. Re-extract and upload the data afresh once the cleaning is concluded.

### Other related references:

For further reference on how to use DWAPI application, refer to the following link.

- i. <https://kenyahmis.org/>

THE END