PEPFAR

U.S. President's Emergency Plan for AIDS Relief

Data for Accountability Transparency and Impact (DATIM)

SIMS 2.0 Data Exchange Exercise Guidance V2

U.S. Department of State – Office of U.S. Global AIDS Coordinator and Health Diplomacy (S/GAC)

CONTENTS

1 BACKGROUND 3
2 DATA EXCHANGE EXERCISE PURPOSE 3
3 SIMS 2.0 PROBLEM STATEMENT 3
4 TIMETABLE 4
5 DATA EXCHANGE EXERCISE PROCESS 5
6 DATIM SIMS 2.0 COMPREHENSIVE FILE FORMAT DETAILS 7
6.1 CSV Data File Specifications 7
6.2 Data Element Naming Convention and Data Types 8
6.3 CSV File Naming Convention 8
6.3.1 CSV File Size *9*
6.4 Sample SIMS 2.0 CSV File 10
Appendix A _TIME Data Element (Optional Format) 11
Appendix B Example SQL Queries 11
B.1 Example SQL Query 1 11
B.2 Example SQL Query 2 12
B.3 Example SQL Query 3 12
B.4 Example SQL Query 4 12
Appendix C: SIMS Data Exchange Exercise Process 13
Appendix D: SIMS 2.0 Import Process Summary 14

1 BACKGROUND

Agencies are required to submit SIMS data to S/GAC via the DATIM system on a quarterly basis. This guide serves as a resource for participating agencies to facilitate the process of submitting their SIMS 2.0 data to S/GAC for import into DATIM. Data intended for import into DATIM must satisfy strict requirements with respect to the format of the data as well the relationship of the data to the current metadata of the system. Data must be valid with respect to the organization units, data elements, and mechanisms.

2 DATA EXCHANGE EXERCISE PURPOSE

The purpose of this exercise is for participating agencies (CDC, DoD, HRSA, PC, and USAID) to verify and validate the script developed by the DATIM team to correct records containing the same OU/IM/Period combination. This will enforce adherence to the org unit/mechanism/period uniqueness constraint required by DATIM.

Note: The script does not alter the *assessment date*, the integrity of which is important for subsequent analyses.

3 SIMS 2.0 PROBLEM STATEMENT

Utilizing the current SIMS aggregate data format, the underlying DHIS2 software upon which DATIM is built does not support import of multiple assessments with the same implementing mechanism (IM), organization unit (e.g. facility, community, or country), and period (set as daily, the lowest possible in DHIS2). The system will only allow a single assessment at a given organization unit on a given IM on a given day. If an additional assessment with this organization unit, IM, period is in an import file, the first assessment will be overwritten in the DHIS2 system.

This was constraint was communicated to the SIMS Interagency Council (IAC) as part of the requirement management phase and the message was re-emphasized in the import guidance materials for SIMS 2.0. Per the '*SIMS DHIS2 Requirements_Packed_10-02-2014*' file dated October 7, 2014.

However participating agencies have flagged this as an issue that can occur for assessments of Service Delivery Points 1 (SDP-1) and Organizational Assessment Points (OAP) that are part of SIMS 2.0.

4 TIMETABLE

Below is a proposed timeline for key items for the exercise.

Item	Who's Responsible	Proposed Timeline
Updated SIMS Guidance* on: changes to the file format the process for this exercise	DATIM Team	11/9/2016
Agency Data Files containing test data	Participating Agencies	1 or 2 weeks after guidance is made available
Proposed dates for the exercise**	Jointly agreed to by the Participating Agencies and DATIM Team	To be determined by participating agencies. Must be completed no later than December 2.
Begin data exchange exercise**	Participating Agencies & DATIM Team	On receipt of Agency Test Data Files. Exercise must be completed no later than December 2.

* Updated SIMS 2.0 guidance contains specifics needed to address issues created and flagged by SIMS IAC indicating that Agency data import files cannot adhere to stated DATIM systems restrictions (on org unit/mechanism/period uniqueness) communicated on October 7, 2014.

** Data exchange exercise dates align with work being done to resolve security issues and requests to Agencies to submit outstanding data.

5 DATA EXCHANGE EXERCISE PROCESS

The table below is provides detailed step by step action items and responsible parties for the upcoming exercise.

	Who's Responsible	Action Item
Step 1	Participating Agencies	Prepare an appropriately formatted file (see sections 6.1 and 6.2)containing sample test data (see section 6.4)
Step 2	Participating Agencies	Name the File correctly (see section 6.3)
Step 3	Participating Agencies	Send the files to the DATIM Team via the DATIM help desk. If the files are large, zip the files and send the zipped version (see section 6.3.1)
Step 4	DATIM Team	Acknowledge file(s) receipt and provide an estimated turnaround time for feedback.
Step 5	DATIM Team	DATIM team will use the R Script on the Agency files to: 1. Check for and resolve duplicates. Duplicate are defined as records with the same OU/IM/Period combination. How the Script Works: The Script will use the *Assessment ID* (in the Comment Field) to identify and differentiate between visits so as to resolve the duplicates. The Script will generate a new Period date and update that field for all records in the file identified as duplicates. In resolving the duplicates, the Script will look across the all Periods in the entire file to transform the duplicates into unique OU/IM/Period combinations; this will ensure no new duplicates are introduced. The resulting file will meet the OU/IM/Period combination uniqueness criteria essential for import into DATIM. Note: **The script does not alter the assessment date, the integrity of which is important for subsequent analyses.** In addition the script will also perform system validations to: 2. Check for Field data types such as: • SIMS scores consist of options, for instance 0-4; a score of 5 for instance, would be considered to be invalid. • For other data types, such as time stamps, strict formatting rules apply. Refer to the SIMS 2.0 Data Import Reference Guide, the PEPFAR Data Exchange Implementer Community –SIMS 2.0 Data Exchange Exercise for

		Facility Module Guide and https://github.com/jason-p-pickering/datim-validation/blob/master/R/checkValueTypeCompliance.Rfor details 3. Check mechanism time frame The data transmitted for a given mechanism should be active. Thus, if data is transmitted for COP 2016 time frame, the mechanism must be active in FACTS for that time period. Seehttps://github.com/jason-p-pickering/datim-validation/blob/master/R/getInvalidMechanisms.R for details. 4. Check mechanism organization unit validity Mechanisms are assigned to a given operating unit. The data's site assignment should fall within the hierarchy of the organization unit. 5. Check that only valid SIMS 2.0 data elements codes are present. See https://github.com/jason-p-pickering/datim-validation/blob/master/R/getInvalidDataElements.R for details.
Step 6	DATIM Team	DATIM Team will provide feedback via the DATIM help desk to the Agency which will include: • A corrected .CSV data file. If duplicates were detected, this file will contain the newly assigned and updated Periods in the Period field on records identified as duplicate. **Note:** Duplicate are defined as records with the same OU/IM/Period combination • A list of any other system validation violations found (see 2-5 above)
Step 7	Participating Agencies	Participating Agencies should: • Review the corrected file to (1) identify records where the Period field has been modified, (2) ensure that the Assessment Date has not been altered, (3) ensure that the corrected file now meets the OU/IM/Period combination uniqueness criteria essential for import into DATIM and (4) check that no other fields within the file have been modified. See Appendix B for example SQL queries. • Correct any system validation violation identified and reported.
	All	After successful completion of the data exchange exercise, Agencies should proceed with the SIMS 2.0 import process outlined in the SIMS 2.0 Data Import Reference Guide (see Appendix C and D)

6 DATIM SIMS 2.0 COMPREHENSIVE FILE FORMAT DETAILS

6.1 CSV Data File Specifications

- Required columns/fields in the CSV file are listed below; the order of the columns/fields MUST be in the same order as shown below.

Dataelement	period	orgunit	categoryoptioncombo	attributeOptionCombo	value	Comment*

Note: the CSV file may or may not include the headers hence they can be omitted from the file.
***The comment field will be stripped out/removed after all duplicates have been resolved. Removal of this field is for the purpose of importing the file into Dev_De or DATIM Production.**

- The Comment field should contain the *Assessment ID.* This will be used to identify and differentiate the visits (i.e. to determine which group of data elements belong with which visit). It is fundamental that each "visit" is clearly grouped with an *Assessment ID* in the Comment field.
- The *Assessment ID* should be unique to each Participating Agencies' file. Agencies can use whatever format they prefer to create this ID. This ID must be unique ONLY within the submitting Agency – there is NO requirement for uniqueness of this assessment ID across agencies.
- Orgunit values should be reported as UIDs (see below).
- Funding mechanism can be reported as either Code or UID.
- Data element can be reported as either Code or UID (refer to code lists for details found at https://datim.zendesk.com/hc/en-us/articles/201767379-DATIM-Data-Exchange-Guide. NOTE: users must be logged into DATIM Support in order to access this password-protected resource);
- Period should be reported using yyyyMMdd format using calendar year: e.g. 20160101 for January 01, 2016, 20160102 for January 02, 2016, etc. *Please note that the period format is different than the "*_Date" field described in section 6.2*.
- Org unit should be reported using UID (refer to code lists for details).
- Category option combo should include the default value "HllvX50cXC0"; this is constant value that should remain the same for all records in the file
- Attribute option combo – the appropriate funding mechanism Code or UID should be reported (refer to code lists for details)
- Value: type of the value is determined by the data element; use the naming convention outlined below for determining types of data that are accepted (see section 6.2)

6.2 Data Element Naming Convention and Data Types

Field Name	Data Type	Length	Valid Values/Format
*_COMM	string, long text	50k characters	
*_SCORE	integer	zeroPositiveInt	
*_RESP	bool		true / false (all lower case)
_CB	bool		true / false (all lower case)
*_PERC	integer	zeroPositiveInt	
*_NA	bool		true/False
*_Date	date		YYYY-MM-DD (2016-01-01 for January 01, 2016). **Note, this is different from the period date format!**
*_Agency	string, long text	50k characters	
*NUM	integer	zeroPositiveInt	
_DEN	integer	integer, zeroPositiveInt	
*_TXT	string, long text	50k characters	See Appendix A for an additional guidance developed for an optional format for this data element. This format is optional.
*_TIME	string, long text	50k characters	
_KP	bool		
_TYPE	integer	zeroPositiveInt	

6.3 CSV File Naming Convention

This section provides guidelines on the naming convention for the files being submitted for this exercise and import.
Naming Convention: agency_module_reportingperiod_versionnumber.csv

agency	Abbreviation for participating agencies: hrsa, dod, usaid, cdc, pc, and cdc
module	SIMS modules include: facility, community, and above site
reporting period	The reporting periods should correspond to PEPFAR Quarters e.g. fy16q2 (all letters should be lower case) refers to Quarter 2 i.e. January – March 2016
verison number	The purpose of having a version number is to assist us identify individual files in the even that multiple files sent to the DATIM Team.

Example: A user is sending a file that contains the data for HRSA's (agency) above site visits (module) for FY16 Q2 (reporting period).
The naming convention will be as follows: hrsa_abovesite_fy16q2_v1.csv
Additional Notes:
- Use the 3 character file extension: .csv
- Do not use special characters, such as, . \ / : * ? " < > |, except for underscores
- Do not use spaces in the file name
- All letters should be lower case

Number of Files per Agency
Each agency should send 1 file for each data set i.e. 1 for facility, 1 for community, and 1 for above site. If an agency does not have data for a particular dataset, they do not need to include a file for that data set.
Example: following the example above for HRSA there should be 3 files in total i.e.
(1) hrsa_ facility _fy16q2_v1.csv
(2) hrsa_ community _fy16q2_v1.csv
(3) hrsa_ above site _fy16q2_v1.csv

6.3.1 CSV File Size

Each individual file type above should not exceed 20MB i.e. the zipped version should not exceed 20MB. Large files should be zipped up; the zipped version should sent to the DATIM Team via the DATIM help desk.

6.4 Sample SIMS 2.0 CSV File

The example below shows sample records with the same OU/IM/Period combination, the use of the [1]comment field to denote the *Assessment IDs* and represent duplicate records. [1]The *Assessment IDs* shown are just example IDs and do not indicate a preferred format; Agencies can use whatever format they prefer to create this ID. [2]Category option combo should include the default value "HllvX50cXC0"; this is constant value that should remain the same for all records in the file. *The comment field will be stripped out/removed after all duplicates have been resolved. Removal of this field is for the purpose of importing the file into Dev_De or DATIM Production.

Dataelement	period	orgunit	Categoryoptioncombo[2]	attributeOptionCombo	value	Comment[1*]
SIMS.CS_ASMT_DATE	20160101	IHuZkWbFwNK	HllvX50cXC0	13559	2016-01-01	ASMT_0001
SIMS.CS_ASMT_ID	20160101	IHuZkWbFwNK	HllvX50cXC0	13559	ASMT_0001	ASMT_0001
SIMS.F_01_01_HIVQMQI_Q1_RESP	20160101	IHuZkWbFwNK	HllvX50cXC0	13559	3	ASMT_0001
SIMS.CS_ASMT_DATE	20160101	IHuZkWbFwNK	HllvX50cXC0	13559	2016-01-01	ASMT_0002
SIMS.CS_ASMT_ID	20160101	IHuZkWbFwNK	HllvX50cXC0	13559	ASMT_0002	ASMT_0002
SIMS.F_01_01_HIVQMQI_Q1_RESP	20160101	IHuZkWbFwNK	HllvX50cXC0	13559	4	ASMT_0002

Appendix A: _TIME Data Element (Optional Format)

This is an optional way of formatting this field; this field can be left blank or formatted in any way the agency prefers however this format must conform to the data field specifications outlined in section 6.2 i.e. must be data type: string, long text and length: 50k characters.

In an effort to maximize the value of _TIME data point (i.e. the ability to use this data point for analytics in future), we suggest that the value of this field follow the rules below:

- Use the Zulu time format i.e. **HH:MM:SS**
- Use an offset in **HH:MM:SS±hhmm** to indicate the UTC time zone; GMT and UTC and not quite the same.
  See http://redirect.state.sbu/?url=http://redirect.state.sbu/?url=http://redirect.state.sbu/?url=http://redirect.state.sbu/?url=http://redirect.state.sbu/?url=http://redirect.state.sbu/?url=https://en.m.wikipedia.org/wiki/Coordinated_Universal_Time for disambiguation.
- All times should be in the **24 hour format**

Appendix B: Example SQL Queries

Below are example scripts to use to verify the corrected file, they use the originally submitted file and the correct validated file. Note: *dbo.SIMS_Sample_File* is the example table/file name of the original file, *dbo.SIMS_Sample_File_Corrected* is the example table/file name of the corrected file that does not contain duplicates.

B.1 Example SQL Query 1

Purpose: To show records with Period data corrected/changed

```
SELECT      dbo.SIMS_Sample_File.period AS SIMS_Sample_File_Period,
dbo.SIMS_Sample_File.Comment AS SIMS_Sample_File_Comment,
                dbo.SIMS_Sample_File_Corrected.period AS SIMS_Sample_File_Corrected_Period,
                dbo.SIMS_Sample_File_Corrected.Comment AS
SIMS_Sample_File_Corrected_Comment
FROM        dbo.SIMS_Sample_File INNER JOIN
                dbo.SIMS_Sample_File_Corrected ON dbo.SIMS_Sample_File.Comment =
dbo.SIMS_Sample_File_Corrected.Comment AND
                dbo.SIMS_Sample_File.Dataelement =
dbo.SIMS_Sample_File_Corrected.Dataelement AND
                dbo.SIMS_Sample_File.orgunit = dbo.SIMS_Sample_File_Corrected.orgunit AND
                dbo.SIMS_Sample_File.attributeOptionCombo =
dbo.SIMS_Sample_File_Corrected.attributeOptionCombo AND
                dbo.SIMS_Sample_File.period <> dbo.SIMS_Sample_File_Corrected.period
GROUP BY dbo.SIMS_Sample_File.period, dbo.SIMS_Sample_File.Comment,
dbo.SIMS_Sample_File_Corrected.period,  dbo.SIMS_Sample_File_Corrected.Comment
```

B.2 Example SQL Query 2

Purpose: To verify that the Assessment Date has not been altered

```
SELECT      *
FROM        dbo.SIMS_Sample_File INNER JOIN
                dbo.SIMS_Sample_File_Corrected ON dbo.SIMS_Sample_File.Comment =
dbo.SIMS_Sample_File_Corrected.Comment AND
                dbo.SIMS_Sample_File.Dataelement =
dbo.SIMS_Sample_File_Corrected.Dataelement
WHERE       (dbo.SIMS_Sample_File.Dataelement = N'SIMS.CS_ASMT_DATE') AND
(dbo.SIMS_Sample_File_Corrected.Dataelement = N'SIMS.CS_ASMT_DATE')
AND dbo.SIMS_Sample_File.value <> dbo.SIMS_Sample_File_Corrected.value
```

B.3 Example SQL Query 3

Purpose: To check if the contents of the Value field has been altered.

```
SELECT      *
FROM        dbo.SIMS_Sample_File INNER JOIN dbo.SIMS_Sample_File_Corrected ON
dbo.SIMS_Sample_File.Comment = dbo.SIMS_Sample_File_Corrected.Comment AND
 dbo.SIMS_Sample_File.attributeOptionCombo =
dbo.SIMS_Sample_File_Corrected.attributeOptionCombo AND
                dbo.SIMS_Sample_File.Categoryoptioncombo =
dbo.SIMS_Sample_File_Corrected.Categoryoptioncombo AND
                dbo.SIMS_Sample_File.orgunit = dbo.SIMS_Sample_File_Corrected.orgunit AND
dbo.SIMS_Sample_File.Dataelement = dbo.SIMS_Sample_File_Corrected.Dataelement
WHERE dbo.SIMS_Sample_File.value <> dbo.SIMS_Sample_File_Corrected.value
```

Note: This query can be modified to check all the other fields by replacing the Value field.

B.4 Example SQL Query 4

Purpose: To check that the corrected file now meets the OU/IM/Period combination uniqueness criteria essential for import into DATIM

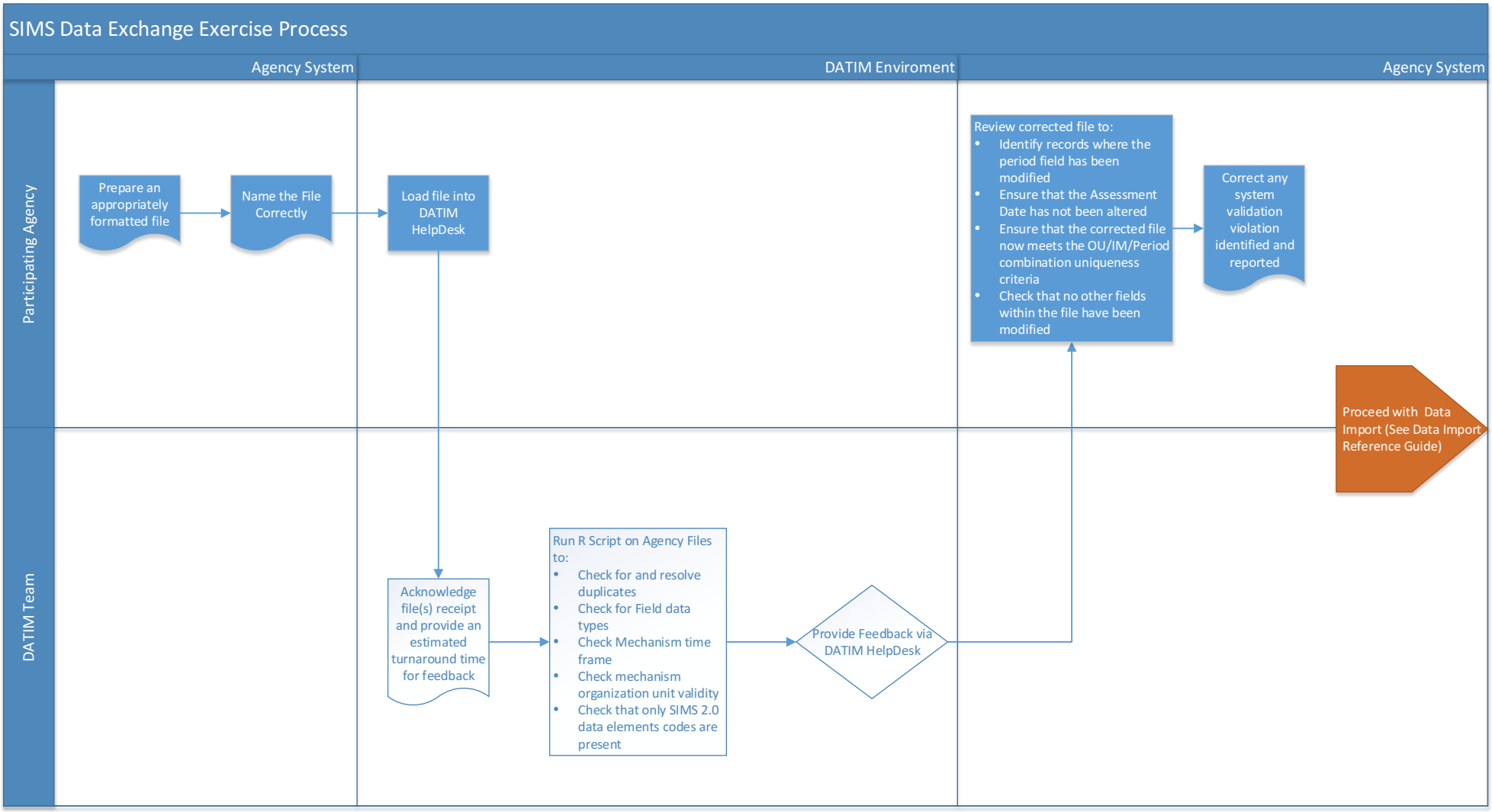
```
SELECT dataelement, period, orgunit, attributeoptioncombo, count (*) as 'count of occurrences'
FROM dbo.SIMS_Sample_File_corrected
GROUP BY dataelement, period, orgunit, attributeoptioncombo
HAVING count (*) > 1
```

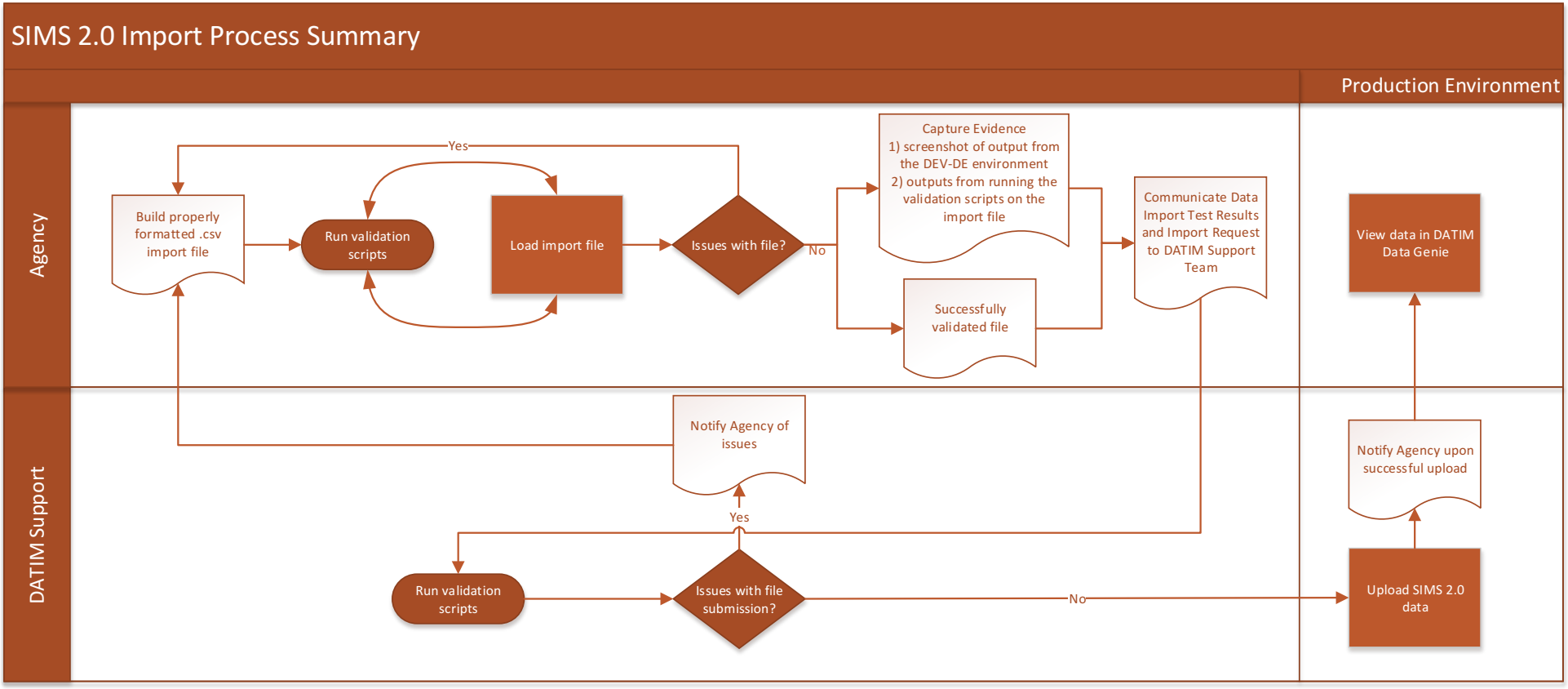
Note: This query will return zero records because no duplicates exist in the corrected file.

```
SELECT dataelement, period, orgunit, attributeoptioncombo, count (*) as 'count of occurrences'
FROM dbo.SIMS_Sample_File
GROUP BY dataelement, period, orgunit, attributeoptioncombo
HAVING count (*) > 1
```

Note: This query will return multiple records as it was executed on the original file containing duplicates.

Appendix C: SIMS Data Exchange Exercise Process



Appendix D: SIMS 2.0 Import Process Summary

This process is for submission to DATIM Production of Actual/Real data. If an agency used Actual /Real data for the above exercise they may not need to go through this entire process- the DATIM Team will provide additional information for this scenario.