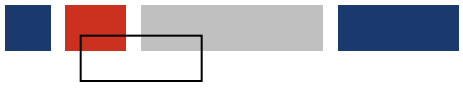
INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

SIMStat Pilot

Instituto Nacional de Estatística
Departamento de Metodologia e Sistemas de Informação
Serviço de Infraestutura Informacional
Sónia Quaresma



September 2014• INE • Portugal

SIMSTAT is a project for micro-data exchange on intra-EU trade

SIMSTAT hub

acting as the central dispatch point for micro-data exchanges between Member States

specified dataset will be used for the exchange between MSs of data files containing Intrastat collected dispatches data in these MSs

Unique Record Identifier (URI)

The URI is a variable string of maximum 30-alphanumeric characters which for each record of files from a MS would be created by that MS and inserted into the file before it was sent to the Hub.
The first two digits will be the country code (from GEONOMENCLATURE) and 3rd – 4th digits would be the year, so e.g. Portuguese would be like "PT14xxxxxxxxxxxxxxxxxxxxxxxxxx" (30-alphanumeric character string).
The URI should be preserved through all the data exchange steps and finally delivered to the recipient.
To avoid problems with cross-country processing, storage and presentation of alphabetical characters in the URIs, only the small and capital letters of the Latin alphabet will be used.
Full uniqueness of a record must be guaranteed across all MSs.

Time of production

The time of production of the record will be presented by a full timestamp (18-alphanumeric character string, not ISO because of minimum length) with the following structure:
YYYYMMDD”T”hhmmss”TimeZone”, e.g. 20140305T232219+02

Sending MS (Reporting MS)

The code list, attached to this coded variable, contains the 2-alphabetic character codes for the list of 28 EU MSs according to the GEONOMENCLATURE.

Receiving MS (MS of Destination)

The code list, attached to this coded variable, contains the 2-alphabetic character codes for the list of 28 EU MSs according to the GEONOMENCLATURE.

Data status flag for receiving MS

The Data status flag for receiving MS is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated
2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Reference period

The reference period is the calendar month of dispatch of the commodity according to the applied national reporting practice.
The reference period is a custom 6-numeric character string: the first four digits = the year; the last two digits = the month, e.g. 201401 for January 2014.

Flow

The code for this variable is a 1-numeric character string “2” for “Dispatches” (as used in DocMet 400 for transmission of ITGS data to Eurostat).

Record action

Record action is a 1-alphabetic character string. The variable indicates the action that should be undertaken with each received record from a SIMSTAT dataset.
A code list containing 2 codes is attached to this variable:

“I” = Insert: The record has to be added to the existing data for the particular reference period;
“D” = Delete: The record with the indicated URI, which should exist in the database of the receiving MS, has to be deleted.
In case an old record has to be updated, a sequence of two logically interconnected records has to be entered in the exchange file:

- the first one with Record action code = “D”, indicating the old record, which should be deleted, lines "D" are sent before lines "I";
- the second one with Record action code “I” indicating the new record to be inserted on the place of the deleted one. The identifier (URI) of the line "I" following a line "D" must be the same for traceability.

The URI is the key in this process, that’s why we proposed the above mentioned changes in Section 2.

ID number of PSI

Optional element

The ID number of PSI is a variable string of maximum 14-alphanumeric characters and for those MSs that would exchange this data element, will be the real VAT Identification Number (VIN) allocated to the trader in the sending MS in accordance with Articles 214 and 215 of Directive 2006/112/EC.

VIN is composed of the 2-alphabetic character codes for the list of 28 EU MSs according to the GEONOMENCLATURE - with the exception of Greece, the code for which is “EL” - and the national VAT number according to the specification of national VAT numbers given in DG TAXUD’s document “VIES – VAT Number Construction Rules Functional Description”, version 16.1 from year 2012.

ID Number of Partner

Mandatory element

ID Number of Partner is a variable string of maximum 14-alphanumeric characters and will be the real or simulated VAT Identification Number (VIN) allocated to the partner trader in receiving MS.

If partner ID is not known and cannot be simulated, a dummy “QV999999999999” may be used.

Data status flag for ID Number of Partner

The data status flag for ID Number of Partner is a 1-alphanumeric character string and indicates the way in which the VIN (VAT Identification Number) of the partner (receiving) trader has been determined. The following code list for this variable is used:

0 – Originally collected variable
1 – estimated with 1:1 relation
2 – error/outlier corrected by PSI
3 – error/outlier auto-corrected
4 – error/outlier not corrected
5 – dummy ID, not intra-EU supply
X – estimated with 1:N relation
Y – dummy ID, no correspondence in VIES

Commodity code

Commodities are reported in accordance to the Combined Nomenclature including codes from chapter 98 and code 9950 00 00.

The codes used are 8-numeric character strings . For Intra-EU estimates, reporting is allowed also at HS6, HS4 and HS2 level (8, 4 or 2-numeric character strings).

A special attention should be made to the commodity codes initiated with zero, which need to be kept.

Data status flag for commodity code

Data status flag for commodity code is a **Mandatory** element and is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Country of origin

Country of origin is an **Optional** element.

The 2-alphabetic character codes according to the GEONOMENCLATURE will be applied when data is known and reported, otherwise the field must be left blank.

Data status flag for country of origin

The Data status flag for country of origin is a 1-numeric character string. The following code list is used:

0 – collected
1 - estimated

Nature of transaction

The code table for “List of Nature of transaction codes” – columns A+B (Regulation (EC) 1982/2004, Annex III) is attached to this variable.

A “0” (zero) is added for column B for column-A codes 6, 7 and 8, or to all column-A codes if data is collected at national level only at 1-digit level (column A).

Data status flag for nature of transaction

The Data status flag for nature of transaction is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated 2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Delivery terms

The code table for “Coding of delivery terms” - Regulation (EC) 1982/2004, Annex IV – Incoterm codes” with the addition of the two new delivery terms from ICC Incoterms2010
“DAP” – Delivered at Place, and
“DAT” – Delivered at Terminal, is attached to this variable and will be applied when data is collected and reported, otherwise the field must be left blank.
Specific terms of delivery, used for national purposes, are not allowed

Data status flag for Delivery terms

The Data status flag for Delivery terms is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated 2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Mode of transport

The Mode of transport is a 1-numeric character string. The code table for “Coding of mode of transport” - Regulation (EC) No 1982/2004, Annex V, is attached to this variable and will be applied when data is available, otherwise the field must be left blank.

Data status flag for mode of transport

Data status flag for mode of transport is a **Mandatory** element when the mode of transport element is reported, or left **Blank** when the mode of transport element is blank.

The Data status flag for mode of transport is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated 2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Quantity in net mass

Quantity in net mass is a positive numeric field expressed in **kg** with **3 decimals** (up to thousandths) with the total length of up to 15 characters which includes Dot as decimal separator and 3 characters for decimals.

Several exception apply.

Data status flag for Quantity in net mass

The Data status flag for Quantity in net mass is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated 2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Quantity in supplementary unit

The Quantity by supplementary unit will be expressed in the particular unit **with 3 decimals (up to thousandths)** with the total length of up to 15 characters which includes Dot as decimal separator and 3 characters for decimals.

Several exception apply.

Data status flag for Quantity in supplementary unit

The Data status flag for Quantity in supplementary unit is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Taxable amount (value)

Taxable amount (value) is a positive numeric value expressed the national currency of the sending MS with 2 decimals (up to hundredths) and the total length of up to 14 characters which includes Dot as decimal separator and 2 characters for decimals.

in Euros for the Member States of the Euro area
in national currency units for the others.

Reporting whole real numbers (without decimal part) is allowed.
Our opinion is that this allowance should be excluded, in order to guarantee better quality of data, avoiding errors (in some cases, by mistake, the decimal part may be consider as integer/real).

Converted taxable amount (converted value)

The Converted taxable amount (converted value) is a positive numeric field and gives the reported values in national currency by the sending MS converted to the national currency of the receiving MS by agreed conversion coefficients stored in the Hub, with the total length of up to 14 characters which includes Dot as decimal separator and 2 characters for decimals.

The following conversions are done by the Hub:
- from non-Euro to Euro;
- from Euro to non-Euro;
- from non-Euro to non-Euro.

If both sending and receiving MSs are members of the Euro area, no conversion is done by the Hub – the values are copied one-to-one.

Data status flag for Taxable amount (value)

The Data status flag for Taxable amount (value) is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated 2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

Statistical value

Statistical value is a positive numeric value expressed the national currency of the sending MS with 2 decimals (up to hundredths) and the total length of up to 14 characters which includes Dot as decimal separator and 2 characters for decimals.

in Euros for the Member States of the Euro area
in national currency units for the others.

Converted statistical value

The Converted statistical value is a positive numeric field and gives the reported values in national currency by the sending MS converted to the national currency of the receiving MS by agreed conversion coefficients stored in the Hub, with the total length of up to 14 characters which includes Dot as decimal separator and 2 characters for decimals.

The following conversions are done by the Hub:
- from non-Euro to Euro;
- from Euro to non-Euro;
- from non-Euro to non-Euro.

If both sending and receiving MSs are members of the Euro area, no conversion is done by the Hub – the values are copied one-to-one.

Data status flag for Statistical value

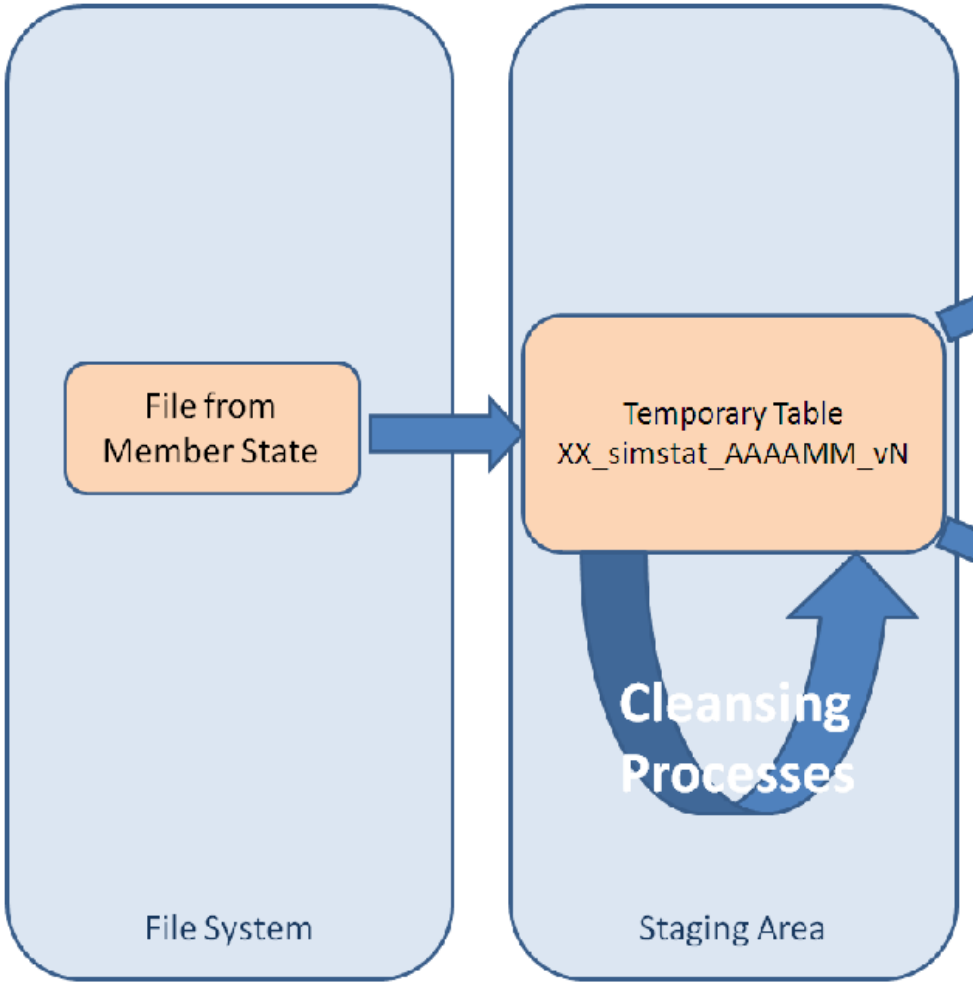
The Data status flag for Statistical value is a 1-numeric character string. The following code list is used:

0 – originally collected variable, not flagged as an outlier
1 – estimated 2 – originally collected variable, flagged as an outlier and corrected after contacting with the PSI
3 – originally collected variable, flagged as an outlier and auto-corrected
4 – originally collected variable flagged as an outlier, but not corrected
5 – originally collected variable, flagged as an outlier, but confirmed after contacting with the PSI
6 – other methods or dummy value

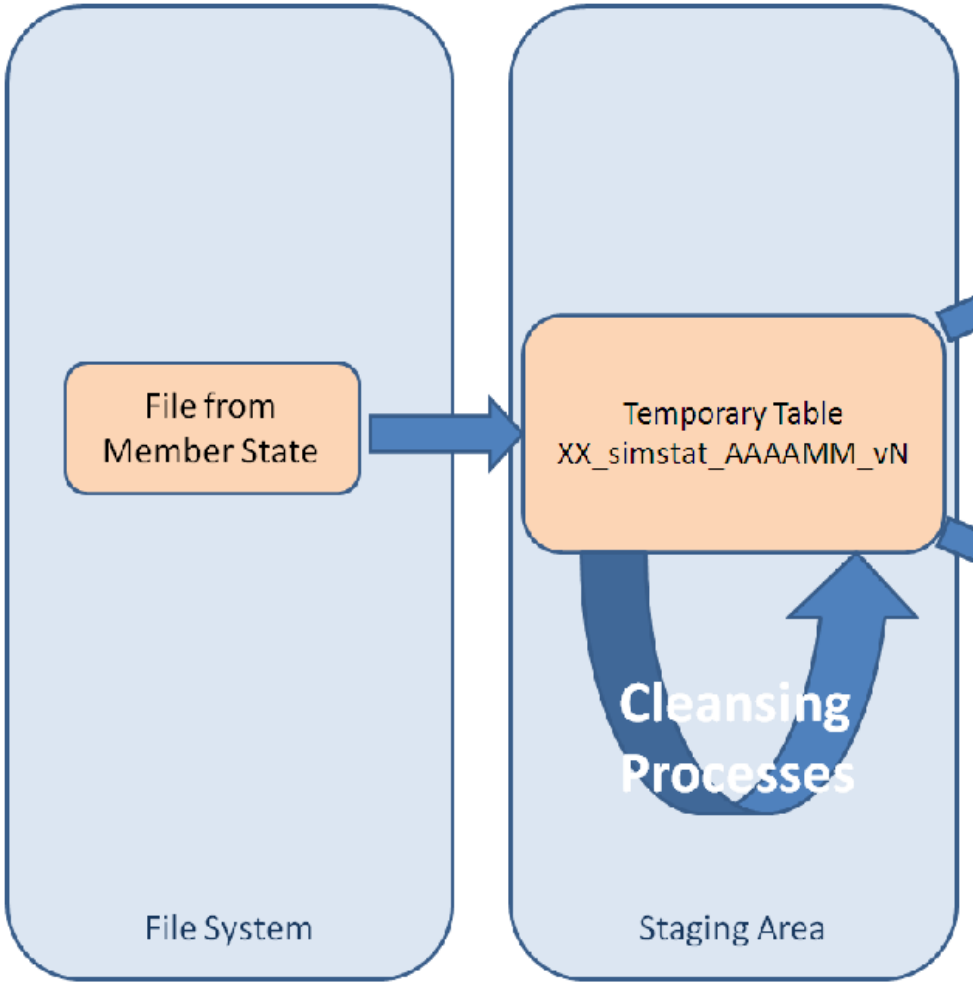
31 Fields in a Dataset


Architecture

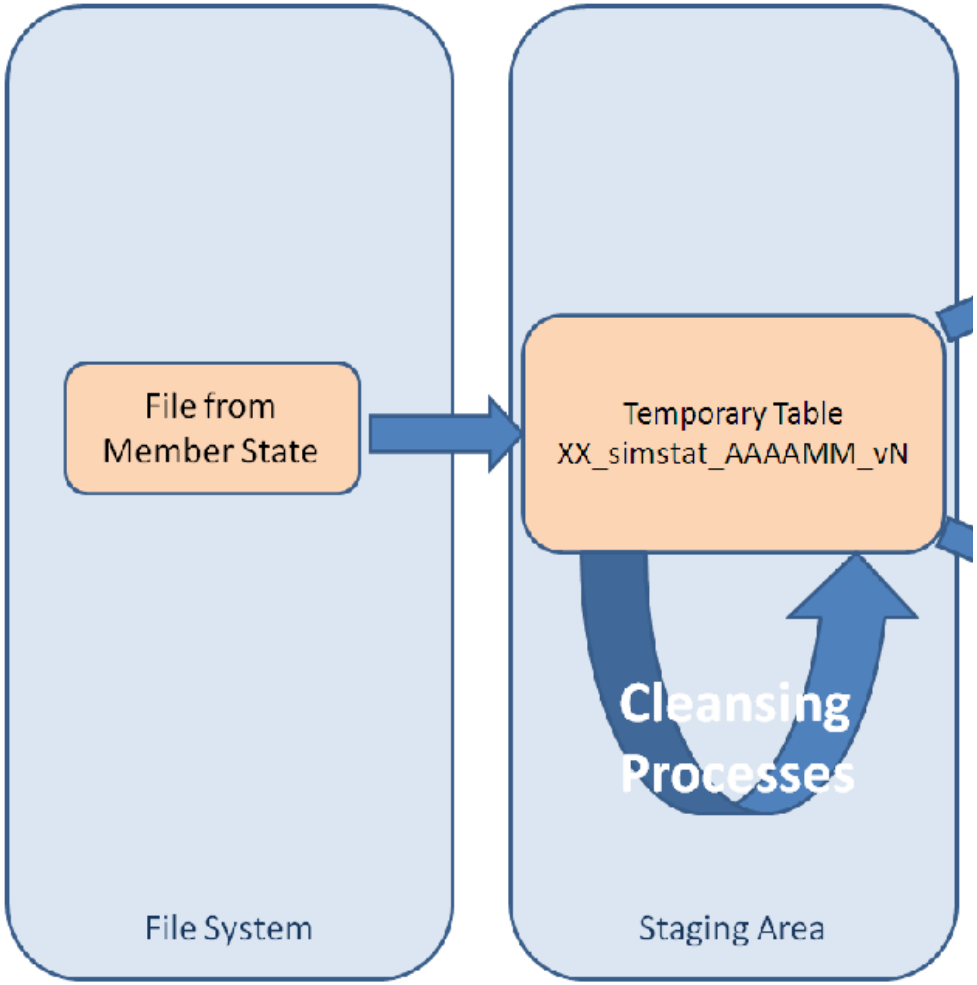
In the SA data is cleansed



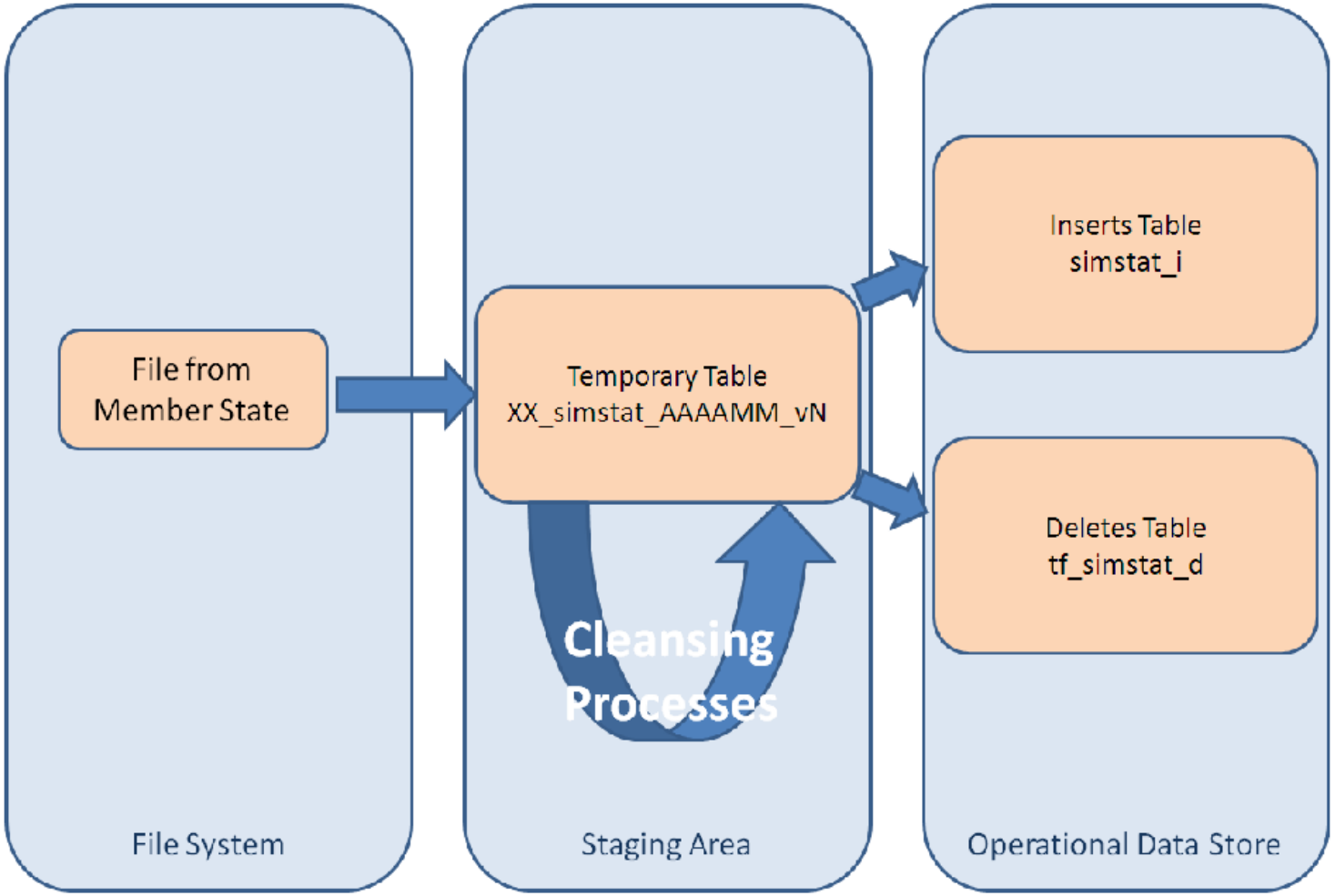
Timestamp format is enforced



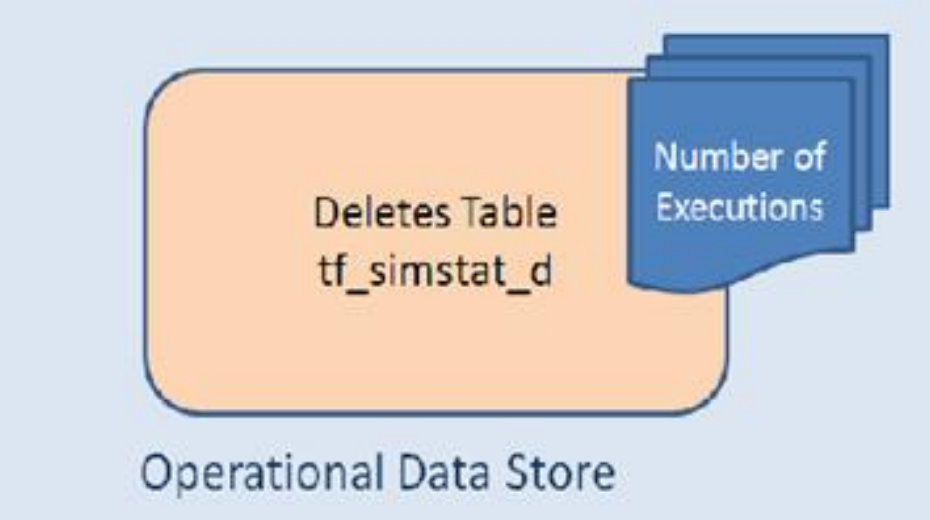
Commodity codes validity is checked



Data is divided – Record Action



Remove Records are inspected for their reference period and time of production



A selection grouping all the records by URI is performed on the table with the delete statements to check the number of times the deletion process should be executed upon the data - *iterations_number*

Unique Record Identifier (URI)

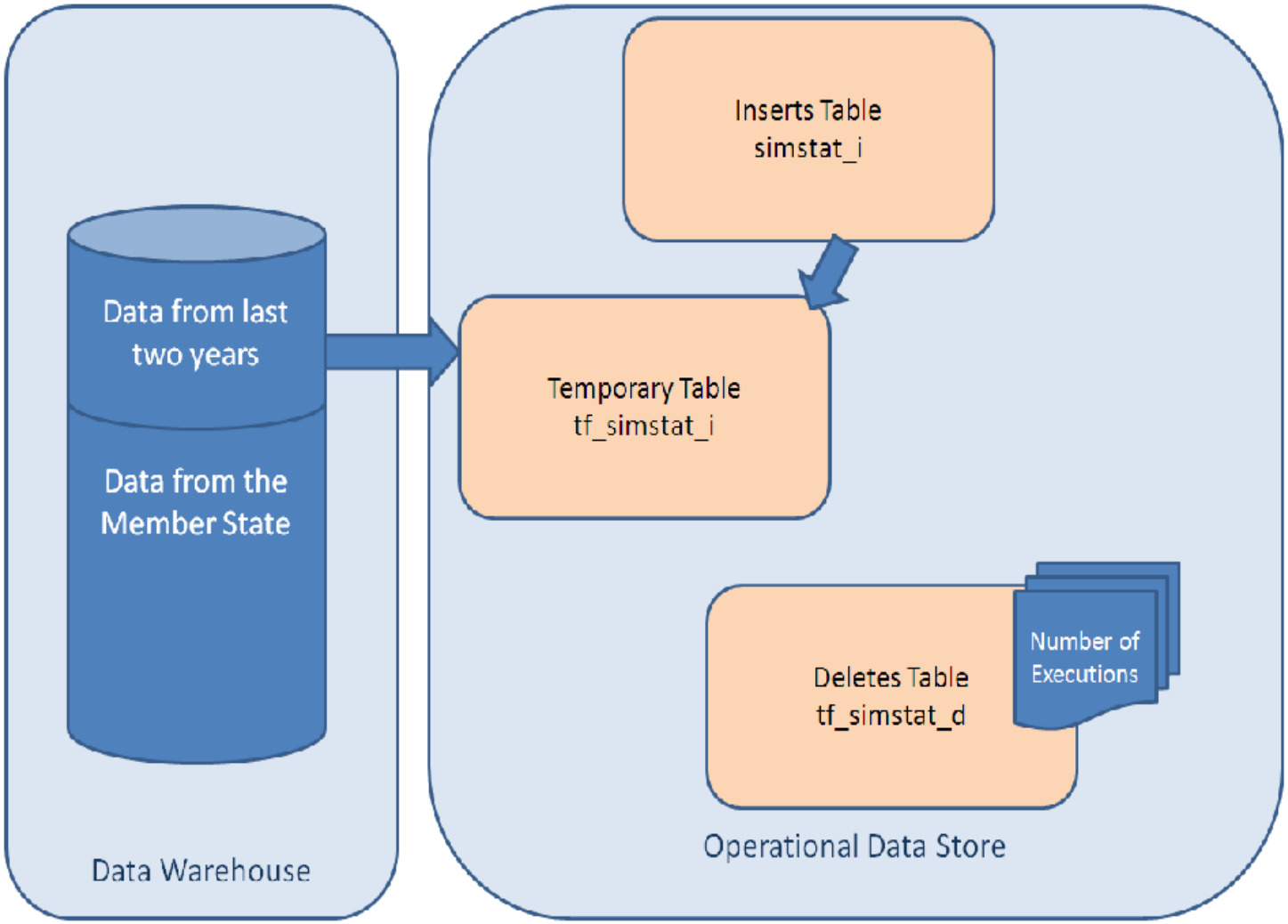
PT14ABH18PTRV358HJA3AQMNB25GDI

Time of production

20140905T232219+01

Record action

D


Inserts Table
simstat_i
Data from last two years
Data from the Member State
Temporary Table
tf_simstat_i
Deletes Table
tf_simstat_d
Number of Executions
Data Warehouse
Operational Data Store

Unique Record Identifier (URI)

PT14ABH18PTRV358HJA3AQMNB25GDI

Time of production

20140905T232219+01

Record action

D

Deletion Process

Deletion Process – table for removal

Table XX_simstat_removal_<E>							
data from record with RA = 'I'				data from record with RA = 'D'			
URI	time_production	sending_ms	…	del_URI	del_time_production	del_sending_ms	del…

Information from the original record in the temporary insert table will be stored side by side with information from the record that issued its deletion

The record with the earliest time of production is selected

Deletion Process – Each iteration

Table XX_simstat_removal_<E>							
data from record with RA = 'I'				data from record with RA = 'D'			
URI	time_production	sending_ms	…	del_URI	del_time_production	del_sending_ms	del…

In each erase iteration the records to be deleted from the target table (tf_simstat_i) are picked not only by their URI but also selecting the records with the earliest time production stamp to prevent that inserts with the same URI performed after the deletion be mistakenly removed.

These records are then entered in the control table XX_simstat_removal_<E> on the leftmost columns.

Deletion Process – Each iteration

Table XX_simstat_removal_<E>							
data from record with RA = 'I'				data from record with RA = 'D'			
URI	time_production	sending_ms	…	del_URI	del_time_production	del_sending_ms	del…

The records which issued the removal from the deletion table tf_simstat_d are also picked by their URI and earliest time production stamp, as a record if their URI is reused can be inserted and deleted several times over.

Their information is then entered in the control table on the rightmost columns.

Checks Performed in each record

Time_Production is earlier that Del_Time_Production
All the other fields have the same value

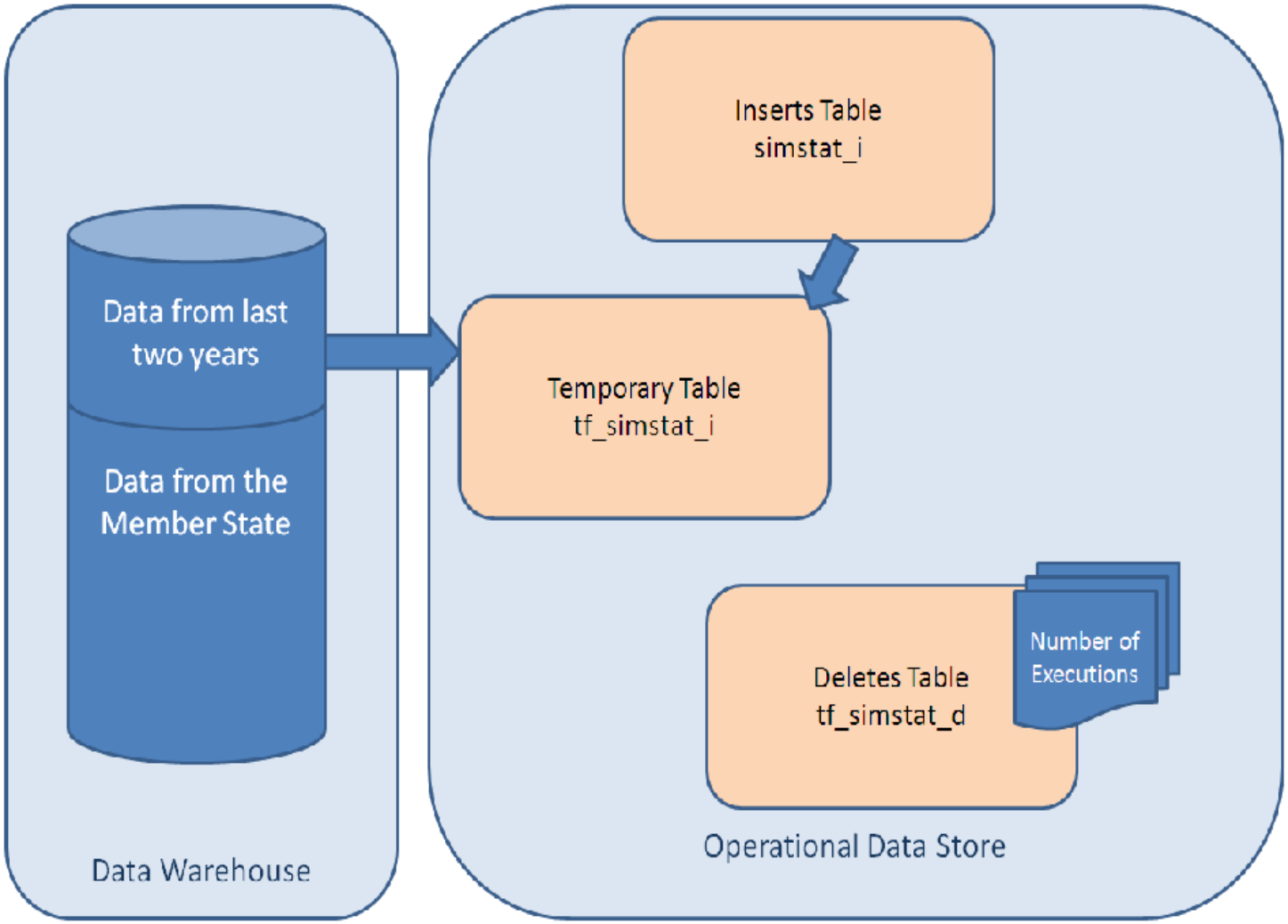
Deletion Process – Each iteration

Table XX_simstat_removal_<E>							
data from record with RA = 'I'				data from record with RA = 'D'			
URI	time_production	sending_ms	…	del_URI	del_time_production	del_sending_ms	del…

The erase process continues by marking all the records on the insert table (tf_simstat_i) which are in the control table XX_simstat_removal_<E> with remove_flag different from N for deletion. And they are then physically removed from the table.

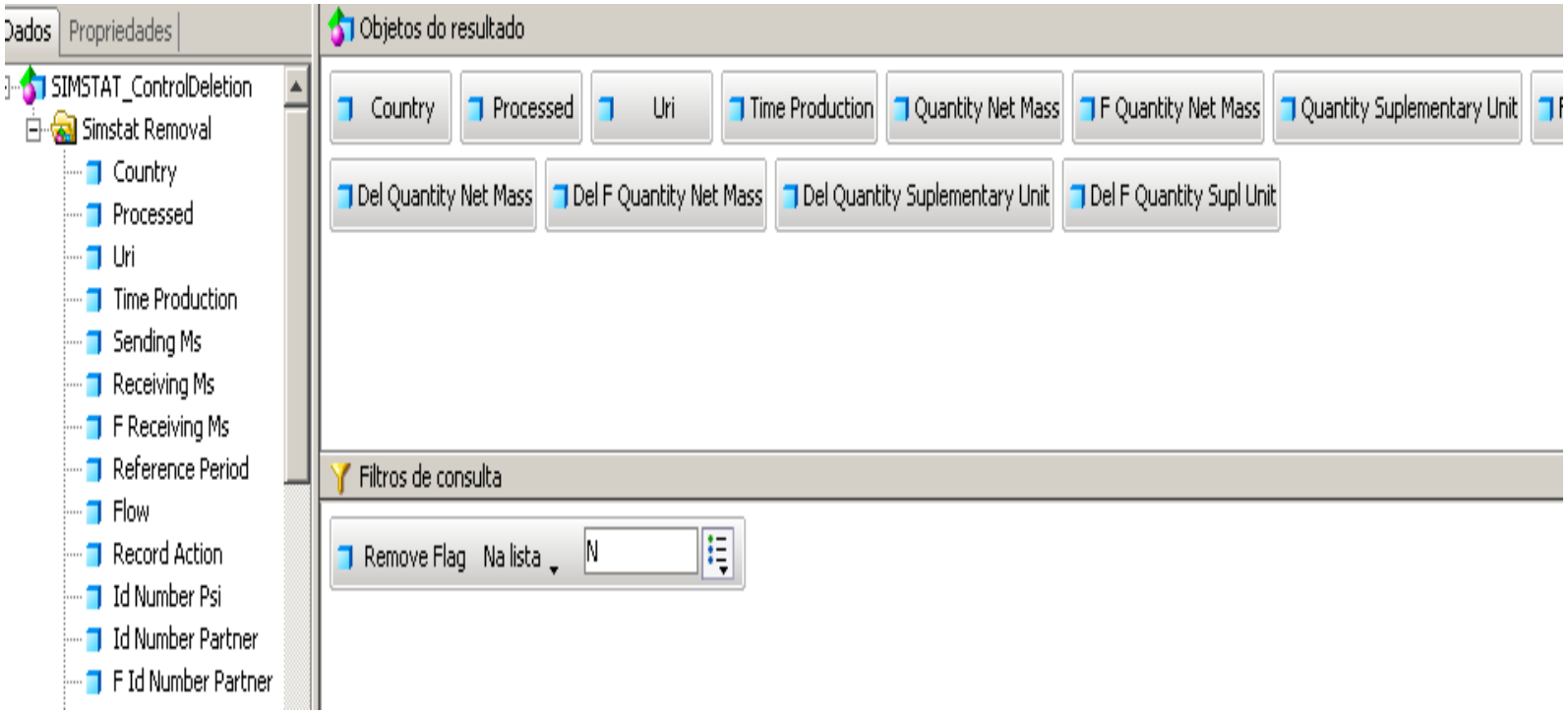
In an analogous process all the records on the delete table (tf_simstat_d) which are in the control table XX_simstat_removal_<E> with remove_flag different from N are marked and physically removed from the table tf_simstat_d.


Deletion is Executed several times


Inserts Table
simstat_i
Data from last two years
Data from the Member State
Temporary Table
tf_simstat_i
Deletes Table
tf_simstat_d
Number of Executions
Data Warehouse
Operational Data Store

Log is used for Quality Process inspection

Table XX_simstat_removal_<E>							
data from record with RA = 'I'				data from record with RA = 'D'			
URI	time_production	sending_ms	…	del_URI	del_time_production	del_sending_ms	del…



Título do relatório

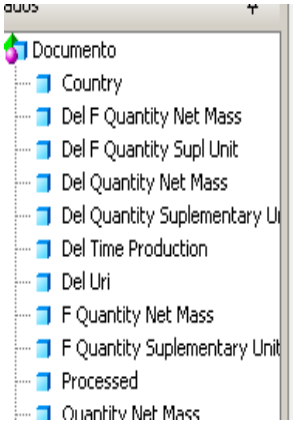
Country	Processed	Uri	Time Production	Del Time Production	Quantity Net Mass	Del Quantity Net M
DE	20140723	DE120000000000(	20120306T224041+01	20121002T095803+01	12	1
DE	20140723	DE120000000000(	20120803T165404+01	20121002T161521+01	261	162
DE	20140723	DE120000000000(	20120803T223403+01	20120815T104433+01	8	28
DE	20140723	DE120000000000(	20120904T134527+01	20121002T161252+01	594	464
DE	20140723	DE120000000000(	20130103T195911+01	20130306T134618+01	57	2


Update processes

Update processes

During the examination of the records on control table XX_simstat_removal_<E> where the fields on the deletion record do not match those previously on the insertion record the analysts may decide to perform an update.

The update is done exchanging the records in the control table.
The field remove_flag on the control table XX_simstat_removal_<E> is then altered to U to indicate an update.



Título do relatório

Country	Processed	Uri	Time Production	Del Time Production	Quantity Net Mass	Del Quantity Net M
DE	20140723	DE120000000000	20120306T224041+01	20121002T095803+01	12	1
DE	20140723	DE120000000000	20120803T165404+01	20121002T161521+01	261	162
DE	20140723	DE120000000000	20120803T223403+01	20120815T104433+01	8	28
DE	20140723	DE120000000000	20120904T134527+01	20121002T161252+01	594	464
DE	20140723	DE120000000000	20130103T195911+01	20130306T134618+01	57	2