



SDMX STATISTICAL GUIDELINES

GUIDELINES FOR CONFIDENTIALITY AND EMBARGO IN SDMX

Version 1.0
15 November 2015

Introduction

These guidelines cover the confidentiality aspects in SDMX data exchange, including embargo scenarios within the scope of a single transmission. The aim is to provide a consistent and practical way to represent these aspects in SDMX artefacts in order to promote cross-domain consistency and harmonise methodology and processes.

The paper presents the use case scenarios related to confidentiality and embargo. Based on the use cases, recommendations are provided on how to represent both elements in the SDMX model.

Data are protected by confidentiality in cases where unauthorised disclosure could be prejudicial or harmful to the interest of the source or other relevant parties.

These guidelines follow the definition of embargo on data in the SDMX Glossary, implying that any data being non-publishable before the expiry of the embargo must become public after expiry of the embargo.

Embargo establishes a relationship between a set of data (e.g. an observation), a date/time and a group of privileged data recipients. Beyond the embargo date/time, it is available to the general public.

Unauthorised disclosure of data that are confidential or under embargo is not permitted. Procedures should be in place to prevent disclosure of these data, including rules applying to staff, aggregation rules when disseminating data, provision of unit records, etc.

There should be a formal agreement between organisations involved in the exchange of confidential data in order to prepare systems and workflows.

Data exchange partners are advised to agree up front on the usage of the embargo mechanism(s) for specific data messages.

Confidentiality Use Cases

This section describes the confidentiality and embargo use cases that are addressed by these guidelines. The use cases and embargo SDMX representations are summarised in annex 1:

A) USE CASE: NON-CONFIDENTIAL DATA

Data is available to the public immediately, data is not confidential and there is no embargo.

The data's CONF_STATUS attribute should be set to "Free (free for publication)".

SDMX representation

CONF_STATUS: F

B) USE CASE: CONFIDENTIAL DATA

One or more observations in the data message are confidential. Embargo does not play a role in this use case. Depending on arrangements between data exchange partners, these data can be made available to privileged data users.

The observation's CONF_STATUS attribute should be coded as something other than F "Free (free for publication)". There are several confidential status; below are some examples of their uses; however users are referred to the CL_CONF_STATUS cross-domain code list¹ for a full list of codes:

- N: Not for publication, restricted for internal use only. Used to denote observations that are restricted for internal use only within organisations. This is often used for embargoed data (see below)
- C: Confidential statistical information (primary confidentiality) due to identifiable respondents
- D: Secondary confidentiality set by the sender, not for publication
- A: Primary confidentiality due to small counts

If confidentiality is used without embargo this is the SDMX coding:

SDMX representation

CONF_STATUS: [C;D;S;A;O;T;G;M]

C) HOW TO ALLOW FORWARDING OF CONFIDENTIAL DATA TO SECONDARY RECIPIENTS

Certain data sent to certain primary recipients are confidential, and the sender explicitly allows these primary recipients to forward the confidential data to a restricted set of secondary recipients.

The observation's CONF_STATUS attribute should be marked as "Not for publication, restricted for internal use only". An additional observation-level attribute: CONF_REDIST, defines the secondary recipient(s) to whom the sender allows the

¹ http://sdmx.org/wp-content/uploads/CL_CONF_STATUS_v1_1_26-6-2014.doc

primary recipient to forward confidential data². See section A) **Use of the CONF_REDIST ATTRIBUTE** for the appropriate coding of this attribute.

The forwarding of confidential data may be combined with embargo information (see below).

SDMX representation

CONF_STATUS: N; **CONF_REDIST** (Observation, Conditional): [Organisation(s)];

D) HOW TO ADD EMBARGOED INFORMATION TO A DATA MESSAGE

Following the definition of embargo, data is confidential to the recipient until a point in time (embargo time) and then becomes public.

Allowing privileged access to embargoed data

If the goal is to allow privileged data recipient's access to several embargoed observations in a data message (message), the embargoed observation's CONF_STATUS attribute should be coded as "Not for publication, restricted for internal use only" with an observation level attribute EMBARGO_TIME (date/time/time zone).

SDMX representation

CONF_STATUS: N; **EMBARGO_TIME** (Observation, Conditional): [timestamp]

Enabling the frontloading of data into systems

If the goal is to allow frontloading of a whole data message into systems so that the data can be made visible to users at the expiry of the embargo date/time, the header section of the message should contain an embargo date/time. This implies that all information in the data message is under the embargo date/time set in the header. The header attribute EmbargoDate with format date/time/time zone indicates until when the whole data message received cannot be shared with any recipient users.

Once the EmbargoDate in the header elapses, each observation's confidentiality status becomes that which is marked in the CONF_STATUS attributes.

Note that the frontloading scenario presumes that all data in the message cannot be viewed before the header EmbargoDate, and that there is no privileged access before this time. However, observations may be marked with CONF_STATUS:N "Not for publication, restricted for internal use only" that may be available for privileged access after the frontloading EmbargoDate elapses.

SDMX Representation

CONF_STATUS: < Set to the required confidentiality status after the embargo time elapses.>; <Header>\<EmbargoDate>: [timestamp]

The two ways of representing embargoed data exist to provide efficiency in the exchange, allow for differentiating between data intended to be frontloaded and data aimed to be provided in advance to a restricted audience, and provide flexibility when few data points need to be embargoed in a large data message. The trade-off is the

² Example: Organisation NSI XX reporting data to Eurostat indicates that Eurostat can forward those data to the ECB, IMF and OECD. More complex use case: The reporting organisation may indicate that Eurostat can forward those data only to the ECB Statistics department but no other organisations and also no other ECB departments.

complication of system implementation to support the two representations of embargo, which has to be done locally on a case-by-case basis.

Additional recommendations and examples

In data flows that feature confidential data, CONF_STATUS is highly recommended to be a mandatory attribute. However, if CONF_STATUS is optional in the DSD and missing from an observation, it is always implied to be “F” (free).

A) USE OF THE CONF_REDIST ATTRIBUTE

The CONF_REDIST attribute defines the secondary recipient(s) to whom the sender allows the primary recipient to forward confidential data. It is recommended to be an optional attribute at observation level. Ideally it should reference a shared code list containing standard organisation codes. To allow several secondary recipients there are these possibilities:

1. Use a compound organisation code following the coding guidelines to build aggregates representing multiple organisations, or;
2. Use several CONF_REDIST attributes to portray the multiple recipients. Each attribute represents one recipient and references the same codelist. This implementation is cleaner than the above point 1, though it should be noted that the minimum number of attributes in the DSD is the maximum number of organisations who could be recipients together at the same time.

If the EMBARGO_TIME and CONF_REDIST attributes are both used:

1. Data is available only to the organisations in CONF_REDIST until EMBARGO_TIME
2. Data is available to the public after EMBARGO_TIME

B) CHANGING THE CONF_STATUS FLAG FOR DATA EMBARGOED AT OBSERVATION LEVEL IN A SINGLE TRANSMISSION

In order for an organisation to automate the publication of data once the embargo time elapses, a system process or mechanism is required to change the confidentiality status of the embargoed observation(s) from “N” to “F”.

Example requiring multiple transmissions of confidential data

The solutions suggested above aim at covering the most common confidentiality and embargo use cases within a single transmission from the primary reporter to the primary recipient. However, for some more complex scenarios it might still be required to make multiple transmissions.

In the case where the recipient of the data cannot unambiguously adjust the code value of the attributes; or if a modification of the CONF_REDIST and CONF_STATUS attributes is deemed risky by either the data provider or data recipient(s), then the data transmission should involve multiple messages.

It is strongly recommended that such cases are specified in an agreement between organisations involved in regular transmissions up front in order to avoid unnecessary delay in data publication or – much worse – confidentiality breaches.

In addition, some business requirements may make it impossible to specify the full timeline in a single message. The example below shows a typical use case that cannot be met by a single transmission. In this example the business requirement is the following:

1. A Euro area country wants to provide national data for calculation of EU aggregates to Eurostat and the ECB. For certain observations only Eurostat and ECB statistical staff (i.e. no economists) should be allowed to access the data and only use it for their statistical production purposes. For reasons of practicality the country does not want to send the data message to ECB and Eurostat in parallel but has an agreement to send data only to Eurostat. The country allows Eurostat to forward the data to ECB-Stat.
2. After a certain period the national data may also be shared with the whole ECB and additionally the IMF and OECD for internal use in the context of international data cooperation agreements. It is however not yet public. At that stage it is known that data will be public in one month from now.
3. Following the statement above, Eurostat, ECB, IMF and OECD may use the data immediately but only publish it in one month.

Since there are multiple organisations involved at different points in time, the attributes cannot anymore be used to simply define an embargo timestamp and a list of organisations. Multiple transmissions are required: in this case at least two. The related data exchange sequence would be the following:

1. NSI sends data to Eurostat and includes CONF_REDIST information as “only forward to ECB_Stat”:

CONF_STATUS:N; CONF_REDIST: ECB-Stat

2. NSI sends data to Eurostat again and includes CONF_REDIST information specifying that data can be shared with ECB, IMF and OECD. Additionally it is indicated that data is public in one month from now:

**CONF_STATUS:N; CONF_REDIST: ECB+IMF+OECD;
EMBARGO_TIME=<T+1 month>**

3. <T+1 month>: After EMBARGO_TIME elapses the data is made public by everybody. Based on the definition of embargo, no transmission is needed at that stage, unless data exchange partners prefer to do so anyway.

Annex 1: SDMX Representation of the confidentiality use cases

Use case	CONF_STATUS (Observation)	Additional attributes	Remarks
Non-confidential data	F		
Confidential data with no embargo	C;D;S;A;O;T;G;M;N		CONF_STATUS will usually be C but may also be D;S;A;O;T;G;M;N depending on the required status and confidentiality reason. See the CL_CONF_STATUS code list for details ³
Forwarding of confidential data	N	CONF_REDIST: (Observation, Conditional)	CONF_REDIST may represent multiple organisations
Embargo: Privileged access	N	EMBARGO_TIME (Observation, Conditional)	Only the observations with an EMBARGO_TIME attribute are embargoed
Embargo: Frontloading	Set to the required confidentiality status after the embargo time elapses	<Header\EmbargoDate>: [timestamp]	There is no EMBARGO_TIME attribute as the whole message is embargoed with no privileged access

³ http://sdmx.org/wp-content/uploads/CL_CONF_STATUS_v1_1_26-6-2014.doc