

**SDMX STATISTICAL GUIDELINES**

GUIDELINES FOR
CONFIDENTIALITY AND EMBARGO IN SDMX

Version 2.0 - 19 January 2018

# Document History

|  |  |  |
| --- | --- | --- |
| Version | Date | Comment |
| 1.0 | 21/8/2015 | Initial version. |
| 2.0 | 7/3/2018 | Replaced the “Embargo: Privileged access” use case confidentiality status to use CONF\_STATUS:E instead of CONF\_STATUS:N. When this guideline is implemented, the CONF\_STATUS:N can no longer be used for this use case (the embargo time is ignored if the CONF\_STATUS is N).Clarified the document text, removed superfluous text.Added use of time zone is recommended. |
|  |  |  |

# Introduction

This paper presents use case scenarios related to confidentiality and embargo in SDMX data exchanges, and provides recommendations on how to represent these elements in the SDMX model. 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.

Confidentiality aims at protecting data from unauthorised disclosure that could be prejudicial or harmful to the interest of the source or other relevant parties.

Embargo means that data may become public only after expiry of a pre-defined date and time.

Embargo establishes a relationship between a set of data (e.g. an observation), a date/time and a group of privileged data recipients.

Disclosure of data marked as confidential or under embargo is not permitted. Procedures should be in place to prevent such disclosure, including rules for staff, aggregation rules when disseminating data, provision of unit records, etc.

There needs to 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.

The embargo CONF\_STATUS value “E” is not recommended for final dissemination to users but only for data exchange.

# 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:

## Use case 1: Non-confidential data

Data is available to the public immediately, meaning that 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
 |

## Use case 2: Confidential data

### Exchange of confidential data without embargo nor forwarding to secondary recipients

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

The observation’s CONF\_STATUS attribute should use a specific code denoting the confidential character of the information. Below are some examples of such confidentiality statuses[[1]](#footnote-1):

* **N**: Not for publication, restricted for internal use only. Used to denote observations that are restricted for internal use only within organisations
* **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

### Forwarding confidential data to secondary recipients

A sender sends confidential data to certain primary recipients, and allows those to forward the confidential data to a restricted and pre-defined 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 primary recipient to forward confidential data[[2]](#footnote-2). See section **Use of the CONF\_REDIST attribute** for the appropriate coding of this attribute.

The forwarding of confidential data is represented as follows in SDMX:

|  |
| --- |
| SDMX representation* **CONF\_STATUS**: N;
* **CONF\_REDIST** (Observation, Conditional): [Organisation(s)];
 |

### Adding embargo information to a data message

Following the definition of embargo, the recipient must keep the data confidential until a pre-defined point in time (embargo) when it becomes public.

Two cases can be distinguished:

* Allowing privileged access to embargoed data
* Enabling the frontloading of data into systems

***Allowing privileged access to embargoed data***

If the goal is to allow the data recipient to have privileged access to embargoed observations in a data message (message), the embargoed observation’s CONF\_STATUS attribute should be coded as “E: Not for publication until the embargo time expires; free for publication after the embargo time expires.” with an observation level attribute EMBARGO\_TIME (date/time/time zone).

|  |
| --- |
| SDMX representation* **CONF\_STATUS**: E;
* **EMBARGO**\_**TIME** (Observation, Conditional): [timestamp]
 |

Including a time zone is strongly recommended and the best case is to use the UTC (Coordinated Universal Time) time standard. However, if no time zone is provided then the time zone of the recipient is assumed.

These two examples represent the same time for a recipient established in the Central European time zone (e.g. Germany, Norway, Gibraltar):

* (Recommended) With UTC indicator: 2017-12-15T14:02:29Z
* With timezone indicator: 2017-12-15T15:02:29+01:00

***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 attribute. 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 this 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 any other confidentiality status that is valid 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 data intended to be frontloaded and data aimed to be provided in advance to a restricted audience, and provide flexibility when few observations need to be embargoed in a large data message. The trade-off is the 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).

### 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 code that represents 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 this will require adding as many attributes to your DSD as there are potential recipients of the redistributed confidential data.

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

|  |
| --- |
| Privileged Access |
| **Use case** | **No forwarding** | **Forwarding** |
| **Embargo** | CONF\_STATUS: EEMBARGO\_TIME | CONF\_STATUS: EEMBARGO\_TIMECONF\_REDIST |
| **No embargo** | CONF\_STATUS: N | CONF\_STATUS:NCONF\_REDIST |

### An example of sending data for privileged access with data forwarding information

This example describes a case where data needs to be embargoed until a certain date and time, and may be sent to certain other organisations in a single transmission without modification of the data or attributes.

This example is based on the exchange of sector accounts statistics within the European statistical system.

* The national statistical institutes send data to Eurostat, and allow the data to be shared with the ECB for statistical coproduction
* The data may only be shared with the public on the next day

**CONF\_STATUS:**E**;**

**CONF\_REDIST:** ECB**;**

**EMBARGO\_TIME=<**T+1 day**,** e.g.2017-12-15T10:00:00Z>

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.

It is strongly recommended that use 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.

**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]](#footnote-3) |
| Forwarding of confidential data | N | CONF\_REDIST: (Observation, Conditional) | CONF\_REDIST may represent multiple organisations |
| **Embargo: Privileged access** | E | EMBARGO\_TIME (Observation, Conditional) | Only the observations with an EMBARGO\_TIME attribute are embargoed. After the embargo time elapses, the data are free for publication (equivalent to F status). |
| **Embargo: Privileged access with forwarding** | E | EMBARGO\_TIME (Observation, Conditional)CONF\_REDIST: (Observation, Conditional) | Only the observations with an EMBARGO\_TIME attribute are embargoed. After the embargo time elapses, the data are free for publication (equivalent to F status).CONF\_REDIST may represent multiple organisations |
| **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.  |

1. For a full list of confidentiality statuses, see <https://sdmx.org/wp-content/uploads/CL_CONF_STATUS_1_2_2018.docx>. [↑](#footnote-ref-1)
2. Example: National statistical institute XX reporting data to Eurostat indicates that Eurostat can forward those data to the ECB, IMF and OECD. More complex use case: The reporting organization specifies that Eurostat can forward those data only to the ECB Statistics Department, thus excluding all other organisations as well as all other ECB departments. [↑](#footnote-ref-2)
3. <https://sdmx.org/wp-content/uploads/CL_CONF_STATUS_1_2_2018.docx> [↑](#footnote-ref-3)