Reference metadata in business statistics

Preface

This chapter provides a description of how data compilers report reference metadata at national level to Eurostat and explains the relationship with quality.

The chapter is part of the online publication European Business Statistics manual, which offers a detailed description of methodologies and background information on how business statistics are produced in the European Statistical System (ESS).

Contents

1. What is metadata?

The global exchange of data is increasing every day. Data dissemination sites run by different organisations are offering more and more services using data exchange standards that support the automation of data extraction. To be able to process all this data efficiently, reference metadata that describe the data should be produced using a harmonised list of statistical concepts within the ESS. Metadata are essential for understanding the data, and allow users to make comparisons between data and assess the quality of data. Metadata can be expressed as text (e.g. descriptions), values (e.g. percentage rates) and codes.

There are different types of metadata. Structural metadata act as identifiers and descriptors of the data, e.g. dimensions of statistical cubes, variables, titles of tables, navigation tree. They must always be associated with the data, otherwise it becomes impossible to identify, retrieve and navigate the data.

Reference metadata are used to describe the data. There can be different description types, for example:

- 'conceptual' metadata, describing the concepts used and their practical implementation;

- 'methodological' metadata, describing methods used for the generation of the data;

and

- 'quality' metadata, describing the different quality dimensions of the resulting statistics.

Reference metadata can be exchanged independent of the data they are related to, but are linked to the data in question.

Legal basis:

The following legal basis has been defined to support the use of reference metadata in the ESS:

- Commission Recommendation 2009/498/EC of 23 June 2009 on reference metadata for the European Statistical System.
- Regulation (EU) 2015/759 of the European Parliament and of the Council of 29 April 2015 amending Regulation (EC) No 223/2009 on European statistics.

2. How to process metadata in the ESS?

Metadata is created based on several steps:

- Mapping existing national reference metadata to the two reporting standards (see below);
- Converting existing national reference metadata files into standards; and
- Inserting the national files into the ESS Metadata Handler (ESS MH).

If a new metadata collection is involved, the national user creates the first metadata file using the ESS MH editing wizard.

Reference metadata and quality reports do not exist for all statistical processes within the ESS, and the existing ones may contain confidential information. As a result, not all quality-related information is made publicly available on the Eurostat website.

In addition, the publication of reference metadata and quality reports depends on statistical domain regulations and is decided on the business side at the level of statistical working groups.

The SDMX Glossary (see link in chapter 6) contains the following definition on reference metadata: 'Metadata describing the contents and the quality of the statistical data'.

Two SDMX-compliant reporting standards are currently used to create, collect and compare national reference metadata in the ESS.

Euro-SDMX Metadata Structure (ESMS)

Is a standard and a user-oriented format for the collection of reference metadata in the ESS. It is based on 18 concepts and enables reference metadata to be provided for a list of concepts derived from the SDMX cross-domain concepts. This standard format is also used for the

reporting of national reference metadata files to Eurostat (Commission Recommendation 2009/498/EC of June 2009). ESMS is SDMX-compliant.

ESS Standard for Quality Reports Structure (ESQRS)

Is a standard and a producer of statistics-oriented format. It is based on 11 concepts and allows users to monitor the quality of the statistics produced, concentrating on the main quality criteria (as mentioned in Article 12 of Regulation (EC) No 223/2009 on European statistics). ESQRS is SDMX-compliant.
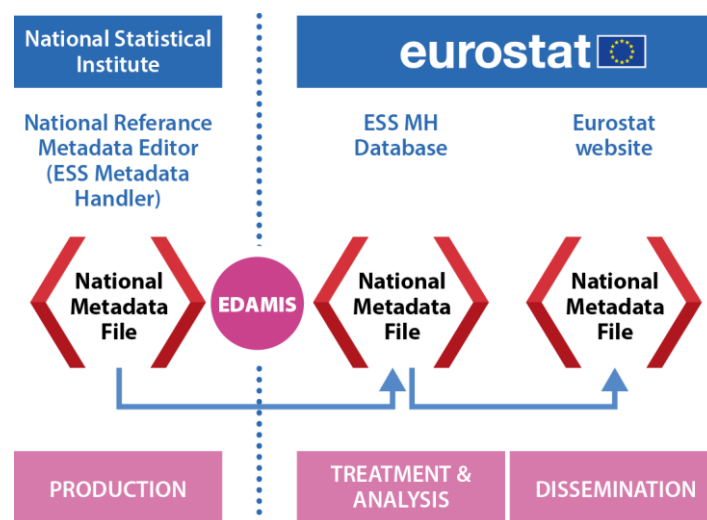
Single Integrated Metadata Structure (SIMS)

SIMS builds on the two above-mentioned reporting structures. SIMS 2.0 with ESMS 2.0 and ESQRS 2.0 were approved by the European Statistical System Committee in November 2015. SIMS will be the standard for quality reporting in accordance with the above-mentioned Article 12 of Regulation (EC) No 223/2009 on European statistics.

Quality is of utmost importance in the world of statistics, and the implementation of SIMS supports quality reporting on European statistics. Producers of official statistics need to ensure that European statistics are developed, produced and disseminated on the basis of uniform standards and harmonised methods. In addition, users of statistics are guaranteed access to appropriate metadata that describe the quality of statistical outputs so that they are able to interpret and use the statistics correctly.

3. The ESS Metadata Handler

The European Statistical System Metadata Handler (ESS MH) is a web-based application that supports the production, exchange and dissemination of reference metadata in the ESS. ESS MH accommodates SDMX-compliant standards for reference metadata (ESMS) and quality reports (ESQRS). It supports the harmonisation of reference metadata and quality reports in the ESS.

The diagram below presents the high-level business process for reporting SDMX-compliant reference metadata and ESS MH usage.

4. Reference metadata on cross-domain coherence in business statistics

A specific part of the national reference metadata is dedicated to *cross-domain coherence* of the dataset(s) with related data.

It describes 'the differences of the statistical outputs in question to other related statistical outputs (incl. main differences in concepts and definitions, statistical unit or object, classification (nomenclature) used, geographical breakdown, reference period, correction methods, etc.). The order of magnitude of the effects of the differences should be assessed as well. For each output the report should contain an assessment of incoherence in terms of possible sources and their impacts.'

In the field of business statistics, coherence metadata with other statistics will address the following issues once the Framework Regulation Integrating Business Statistics is adopted:

- Coherence with the Business Register (e.g. are population, sampling frame and statistical units taken from the national business register? Are the sampling frames for all FRIBS statistics taken on the same date?);
- Coherence with other datasets within the same topic and subject area (e.g. is the number of importing and exporting enterprises consistent with the number of active enterprises? Is the R&D personnel in foreign-controlled enterprises consistent with the R&D personnel total?). For a detailed overview of subjects and topics, see here;
- Coherence with similar datasets of other subject areas of business statistics (e.g. country versus regional turnover data, annual versus infra-annual turnover data; number of employees and self-employed persons collected for ICT versus number of employees and self-employed in active enterprises); and
- Coherence with national accounts (e.g. on investments or labour-related variables) and, where applicable, balance of payments.

5. See also
Overview of methodologies of European business statistics: EBS manual
Legal provisions related to Reference metadata can be found in the following overview

6. Further Eurostat information

Eurostat website:

http://ec.europa.eu/eurostat

ESS Metadata Handler:

https://webgate.ec.europa.eu/estat/spe/metaconv/

Metadata and quality:

http://ec.europa.eu/eurostat/data/metadata

Quality, SIMS and Quality Handbook:

http://ec.europa.eu/eurostat/web/quality/quality-reporting

http://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf/47c0b80d-0e19-4777-8f9e-28f89f82ce18

http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf

7. External links

Information about the Standard for Data and Metadata eXchange (SDMX):

https://sdmx.org/

8. Contacts:

Eurostat, Directorate B (Methodology; corporate statistical and IT services)

ESS Metadata Handler support: ESTAT-METADATA@ec.europa.eu