

# Reference metadata: a step towards greater accessibility and clarity of statistical data

Ana Isabel Sánchez-Luengo Murcia

Asunción Piñán Gaviria

Providencia Núñez González

Luis López Coronado

National Institute of Spain (INE)

## 1. Abstract

The European Statistics Code of Practice lists fifteen European statistics principles seeking to ensure a commitment to quality in the institutional environment, statistical processes and statistical output. Particularly, the 15th principle on accessibility and clarity states that statistics must be supported by metadata for a better and clearer understanding of data. In the National Statistics Institute (INE) of Spain, reference metadata is an integral part of the metadata system, and is also considered to be an essential element in the dissemination of statistical information. The aim of this paper is to present the INE's reference metadata project, its development from its design to its implementation phase, and its convergence with other projects such as the development of the user-oriented quality reports, with the corresponding quality indicators. In describing the INE experience, the problems encountered and progress achieved will be addressed. Finally, an overview of the future projects that the INE expects to carry out on its way to the implementation of an institutional metadata system is shown.

## 2. Reference metadata systems

On February 2005 the Statistical Program Committee adopted The European Statistics Code of Practice which was revised in 2011 and adopted again by the European Statistical System Committee. Principle 15 of this code, regarding Accessibility and Clarity highlights that the data must be accompanied by the metadata, which shall be generated in compliance with standardised metadata systems.

Reference metadata is an important part of the metadata system. It shall be understood as the information that accompanies statistical operations and allows knowing the contents, methodologies and aspects of quality associated with each statistical operation in a systematic, homogeneous and structured way.

In 2009 the European Commission approved the recommendation<sup>1</sup> regarding Euro-SDMX Metadata Structure (ESMS). The annex of this recommendation includes the structure of the reference metadata, which consists in 21 main concepts that are collected using 61 sub-concepts. These 21 concepts can be grouped into 5 large theme groups; presentation of the data, governance, quality, dissemination and statistical production process.

### **3. Standardised methodologies project in the INE**

The standardised methodologies compilation project based on reference metadata (Standardised Methodological Report), started in the year 2010 with the purpose of compiling a metadata system that was integrated and convergent with the European ESMS project. To this end, a multidisciplinary working group was created, which assessed the existent metadata systems.

In Spain, the first record regarding a normalised reference metadata system was the Inventory of Statistical Operations of the State Administration (ISO). Its objective is to make known the way in which statistics are carried out, for coordination and planning purposes, given that it is the starting point of the process of drawing up the national statistical plan. The first ISO was published in 1990 and it has been permanently updated since then. It may be accessed via the INE website<sup>2</sup> since the end of 1999.

The second record was the Special Data Dissemination Standard (SDDS). Spain complies with this standard since it was established in 1996 by the IMF. The SDDS was established to guide members that have, or might seek, access to international capital markets in the provision of their economic and financial data to the public.

---

<sup>1</sup> <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:168:0050:0055:EN:PDF>

<sup>2</sup>

<http://www.ine.es/ss/Satellite?c=Page&p=1254735033477&pagename=IOEhist%2FIOEhistLayout&cid=1254735033477&L=1>

The purpose of these projects are different from the purpose of standardised methodological reports. The latter intends to accompany the data guaranteeing they can be accessed in a clear and normalised way.

This end was also pursued by other examples of reference metadata previous developed by INE, as the methodologies disseminated on the web page. However, these methodologies lacked homogeneity and harmonisation, both in content and extension. Moreover, there was no coordination in its elaboration, which depended on the discretion of each statistical production unit.

The model for reference metadata that the working group was considered most appropriate was the one suggested by Eurostat in its recommendation (ESMS). The ESMS file describes the production, exchange and dissemination of statistics, as well as the data quality and the statistical business process with the advantage of being a normalised metadata system at European level.

Although the most easy and pragmatic solution was to adopt the structured scheme defined in the ESMS, given that all member states have to comply with it, INE project differs in one relevant aspect, that is the unit of reference. The reference unit at national level is the statistic (or what we called "statistical operation") defined as the set of activities, including the preparatory ones, that lead to obtaining and/or disseminating statistical results regarding a particular sector, topic or territory. For Eurostat, however, the unit of reference is the statistical domain<sup>3</sup> as defined by the United Nations and, in some cases, certain indicators and products.

This leads to diverse situations. The statistical operation and the domain coincide in most of the cases, as for example in the LFS (Labour Force Survey). However, in specific cases as the Short-Term Statistics there are not coincident. For a short-term survey as Retail trade indices, nationally a unique methodological report is broadcast, while at European level four ESMS reports are generated, one for each indicator: turnover, persons employed, hours worked and wages and salaries.

---

3

[http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP\\_GLOSSARY\\_NOM\\_DTL\\_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=21227103&RdoSearch=CONTAIN&TxtSearch=domain&CboTheme=&IsTer=&IntCurrentPage=1&ter\\_valid=0](http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GLOSSARY_NOM_DTL_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=21227103&RdoSearch=CONTAIN&TxtSearch=domain&CboTheme=&IsTer=&IntCurrentPage=1&ter_valid=0)

## **4. Two relevant aspects for users: Quality and governance**

Some very important aspects in the production of statistical information and particularly in decision-making are quality management, regulatory aspects and the different institutional policies. These aspects are considered in the methodological reports as a relevant part for the users, also providing information regarding data production costs.

### ***4.1 Quality***

When the reference metadata project started in the INE, the only quality reports that were compiled were those corresponding to statistics subject to this requirement by Eurostat regulation. These reports, which were producer-oriented, did not have a homogeneous structure due to the fact that their structure was fixed in the respective working groups. They were sent to Eurostat and were not published on the INE website. There was therefore a noticeable gap in the communication to users regarding the quality of data disseminated by the INE.

The INE, in its aim of removing the gap in quality communication to users, within its target of implementing a Quality Assurance Framework in the organization, launched a project to gather annually, from the production units, the quality indicators group formerly called "Barometer" and subsequently called "Priority Quality Indicators", for all of the statistical operations they were responsible for.

As the ESMS structure contains a specific group of quality metadata that constitute the user oriented quality report, when the metadata project started in the year 2010, the existing synergy between these two projects became evident. Therefore, the reference metadata file met the need for the INE to systematically implement the compilation of user-oriented quality reports.

When the "Standardised Methodological Reports" started to be compiled, both projects became integrated, including these quality indicators in the ESMS format. This allowed users to access to a very complete quality information in this report.

In this way, with the compilation of the "Standardised Methodological Report" (ESMS), the reference metadata for all the statistical operations published by the INE became available, as well as its user-oriented quality report. With the publication of this report on the INE website, compliance of principle 15 of the Code of Practice is guaranteed

since all information, data and metadata compiled by INE may be accessed in a clear and understandable way.

## ***4.2 Governance***

The INE guarantees coordination, homogenisation and integration of statistics by using the information in the reference metadata reports. Those responsible for decision-making are guaranteed clear and fast access to the updated methodology for each statistical operation.

Since the statistical operation is the unit of reference, convergence is established between the information provided by ISO and that provided by the methodological reports, in concepts such as: objective of the survey, study variables and classification, statistical units the data refer to, way of compiling the statistical operation, administrative sources used, regulation aspects, etc.

Moreover, this information is made available to all users guaranteeing information quality and clarity, with solid and independent methodologies. With information such as: costs and burden on respondents, confidentiality policy, dissemination of data and access to microdata, information exchange agreements, etc.

## **5. Information production process; GSBPM**

On 8 March 2012, the Board of Directors of the INE approved to promote the use of the GSBPM<sup>4</sup> (Generic Statistical Business Process Model) as the language to describe the production of Statistical Operations. This was done on the one hand to guarantee that any statistical operation describes its process in a coherent and normalised way and on the other hand to identify synergies between processes, within the organisation as well as with other organisations.

### ***5.1 Reference metadata to generally describe the statistical process***

Approval and distribution of GSBPM version 4.0 within the United Nations to the international community took place in April 2009. This meant a great advance in metadata systems, which are understood as an essential part of the statistical production

---

<sup>4</sup> <http://www1.unece.org/stat/platform/display/GSBPM/Generic+Statistical+Business+Process+Model>

process. GSBPM version 5.0 approved in December 2013 includes quality and metadata management as an important part of the process. This is considered in the methodological reports as seen in section 4.1.

Reference metadata try to reflect part of the statistical process, specially focusing on the dissemination, data collection, compilation, validation and adjustment stages. Taking into account that they are reports regarding the product and not the process, and that one of its purposes is to allow the users to access the information as clearly as possible. Despite the fact that they are a first approach to the process, they do not cover it completely. They do not include for example phase 1 of need specification or the information regarding how the process is carried out, for example data validation. It does not include the objects considered in GSIM (Generic Statistical Information Model)<sup>5</sup> either.

## ***5.2 Metadata as the driving force of normalisation***

Reference metadata will be used as a tool for normalisation. When compiling new information it will allow consulting variables, definitions and classifications used by other statistical operations as well as production methods.

It will also include metadata regarding processes that are already standardised in the INE. For example, in March 2013 the INE approved the Standard for adjusting seasonal and calendar effects in short-term series. The metadata regarding this process may be found in section 20.6 regarding Adjustment, of standardised reports for statistical operations such as Services Sector Activity Indicators.

## **6. Implementation of the reference metadata project**

An essential part of the project is its execution, which differs from the analysis and modeling. These two stages took place between 2010 and 2012. The project was executed in 2013 and the methodological reports for all statistical operations of the INE were completed and disseminated at the end of this year.

### ***6.1 Implementation***

---

<sup>5</sup><http://www1.unece.org/stat/platform/display/gsim/Generic+Statistical+Information+Model>

On December 13 2011, the Board of Directors of the INE approved the Standard<sup>6</sup> on methodological information for web dissemination. Implementation works consisted in several stages which gave the following results.

For analysis and modeling works it was decided to adopt the ESMS as a model, and the statistical operation as the analysis unit, as well as to develop a web application in Spanish that allowed completing and disseminating information for national Statistical Operations

Execution works were developed during the year 2013. The last test stage of the application was developed and started during the second quarter of 2012 and at the beginning of 2013 it became operational. The taken on commitment was to compile a standardised methodological report following the ESMS standard for each and every one of the statistical operations for which data was published on the INE website.

## ***6.2 Management***

One of the main challenges that faced the project was to ensure that the development of reference metadata be perceived in the organisation as part of the INE's corporate strategy, with benefits and implications for the INE as a whole. Thus the workload for the production units could be taken in a more positive way, making them aware of the advantages of having a usable, comparable, structured and homogeneous metadata information for all statistical operations.

Therefore, in order to facilitate the management of standardised methodological reports, different actions were carried out. These actions were aimed at disseminating the objectives of the project and achieving a favourable climate among all the units in charge of compiling the different statistical operations. During the implementation phase, two training courses and two technical sessions were organised with the attendance of more than 150 members of the staff.

Among these actions it is important to highlight the creation of a multidisciplinary monitoring group integrated by the Coordination, Dissemination, Quality and Standard units. The monitoring group manages the information, dissemination and unification of

---

<sup>6</sup> [http://www.ine.es/en/clasifi/estandar\\_informe\\_metodo\\_en.pdf](http://www.ine.es/en/clasifi/estandar_informe_metodo_en.pdf)

criteria for the compilation of the contents of the different concepts, resolution of doubts and methodological queries, user support in the resolution of problems that are detected during the compilation of their reports by means of computer application, etc.

One of the tasks that were carried out and had excellent results was the customised monitoring of each one of the statistical operations. Monitoring started with an explanatory initial contact in a face-to face meeting with the persons responsible of the operation, so that with enough time before the data was published, the objectives were made known and the appropriate completion criteria were established so that the requested information was understandable, its collection was feasible and possible during the proposed time interval.

To this end, personal, telephone or e-mail assistance was given so as to ensure that the data that was published following the calendar established by the INE included the publishing of the reference metadata report on its corresponding date. There have currently been over 100 published standardised methodological reports.

### ***6.3. Future work***

There are three lines of work still open. On the one hand, the improvement of the information in the methodological reports, and on the other hand, the improvement of IT applications and convergence with other national and international projects. Lastly, despite the main objective of this project was its application to the statistical production of the INE, the first steps have already been taken in order to introduce it in all statistical production carried out by all ministerial departments for state purposes.

The quality of the information in the reference metadata files is very important in each statistical operation as well as among the different domains. The monitoring group is in charge of making this information coherent among the different domains and guarantee that all information is appropriately completed within a statistical operation.

Moreover, works are being carried out so as to guarantee coherence with the information included in the different IT applications available to users. For example right now the information regarding concepts included in the item 3.4 of the ESMS is aligned with the glossary of concepts disseminated on the web page through the



software DEFine<sup>7</sup>. During 2014 we expect to do the same in other items as statistical units (item 3.5) and classifications (item 3.2).

An integrated metadata system is being compiled for the convergence of metadata projects. The system will include a single entry point for quality, process and reference metadata. The idea is that for each statistical operation, to get this information in a combined way and that it is only completed once. This way avoiding duplication and making the information more efficient.

## **7. Lessons learned**

The execution process of the project lasted one year. It was well-received by the production units and cooperation was broad. It does however require a multidisciplinary working group that coordinates the works and guarantees quality and coherence of the information that is introduced and keeps this information updated. Online dissemination of the reports was essential to enhance the commitment of the production units. It was also very important to have our own metadata editor for the completion of the information.

Convergence with other projects at a national and international level has been crucial, although some problems have been encountered such as the different units of reference (statistical operation compared with domain), the problem of language or interoperability between the different applications at a national and international level.

## **8. References**

- [1] Sánchez-Luengo A. (2013) Development of metadata in the National Statistical Institute (INE) of Spain. Joint UNECE/ Eurostat/ OCDE work session on statistical metadata (METIS) Geneva May 2013
- [2] Blanco Frias M. (2006) Statistical Metadata Model Developed at Spain: Current and future use and applications. Joint UNECE/ Eurostat/ OCDE work session on statistical metadata (METIS) Geneva May 2013

---

<sup>7</sup> <http://www.ine.es/DEFine/?L=1>

[3] Pedro Revilla & Asuncion Piñán (2012). Implementing a Quality Assurance Framework based on the Code of Practice at the INE Spain. European Conference on Quality in official Statistics 2012.

[4] UNECE Secretariat (2009). Generic Statistical Business Process Model v 4.0. Joint UNECE/ Eurostat/ OCDE work session on statistical metadata (METIS), Luxemburgo April 2009.

[5] COMMISSION RECOMMENDATION of 23 June 2009 on reference metadata for the European Statistical System

[6] Eurostat. 'The production method of EU statistics: a vision for the next decade', COM(2009) 404 final

[7] Steering Group on Management of Statistical information Systems (2009). Challenges in managing information systems in statistical organizations. Meeting on the Management of Statistical Information Systems, Oslo May 2009.

[8] Zeila K. (2009). Metadata Driven Integrated Statistical Data Management System. Conferencia 'Modernisation of Statistics production', Stockholm November 2009.