

Metadata Life Cycle – Statistics Portugal

Isabel Morgado - Statistics Portugal

A major coordination tool to Statistics Portugal is its integrated statistical metadata system. It supports both statistical production and dissemination of statistical data, and is composed by four main subsystems: Concepts, Classifications, Variables and Statistical Sources, with close interrelationships.

Concepts and definitions used in surveys of the National Statistical System are stored in a database, classified by statistical domain. These concepts are afterwards analysed and approved by the Statistical Council. Once approved their use is mandatory.

The classifications system is based on a classifications database. National, European and International classifications are catalogued and most of them are stored in the database or referenced by a link to their original sites. Classifications are stored in a standardized way and can be downloaded. Each classification is characterized by its own metadata attributes. The main classifications used in the NSS are approved by the Statistical Council. Observation variables from questionnaires and dissemination variables - available in our portal - are stored in a Variables Registry. Names of variables are standardized using a naming convention. Value domains of variables are considered as classifications and are stored in the classifications database.

Methodologies used in surveys are organized in a standard format approved by the Statistical Council. This standard format comprises General Attributes with administrative metadata about the survey, such as name, objectives, legal base, type of statistical operation and others; Methodologies applied like sampling, imputation, estimation and quality evaluation methods, are described. There are also links to the concepts and definitions, classifications, variables and questionnaires used. A last chapter has a list of related documents with links to the documents themselves.

Survey managers are responsible for the production of all the documentation according to the standard format. A Metadata Unit, responsible for the central coordination of this system, has conceived and implemented it in accordance with survey managers, and deals with harmonization issues.

Statistical Council, which assure the coordination of the National Statistical System, approve concepts, classifications and other technical coordination tools. This paper aims to describe the workflow established for these approvals and the interrelationship between the survey life cycle and the metadata life cycle.

isabel.morgado@ine.pt

Register of Technical Projects and Questionnaires in the Basque Statistical Organisation

Marina Ayestarán, Elena Goni, Cristina Prado - Basque Statistics Office-Eustat, Spain

The European Statistics Code of Practice stipulates in his 4th principle that there is to be a "Quality commitment": all ESS members commit themselves to work and cooperate according to the principles fixed in the 'Quality declaration of the European statistical system'.

In the final report drafted by the Leadership Expert Group (LEG) on Quality one chapter is devoted to documentation, as an important aspect of quality. In that report, we can read that: "Documentation activity is regarded as a tool for ensuring and improving quality, on one hand and for better using and understanding data, on the other hand."

The Law on Statistics of the Autonomous Community of the Basque Country establishes that it is compulsory to have a technical project approved by the Basque Statistical Office-Eustat for any statistical operation included in the four-year statistical plans and in the annual statistical plans. The questionnaires and maintaining a questionnaire Register also need to be approved.

The legislation developed in the frame of the Law on Statistics, by means of the April 2005 Decree on Technical Projects, established the approval procedure and creates the Register of Technical Projects and Questionnaires.

The technical project is the document in which the implementation procedures of statistical operations are systematised in all their phases and is to have the following contents to develop:

1. Definition, objectives and dissemination,
2. Methodological design,
3. Implementation procedures,
4. Functions and responsibilities,
5. Quality Control and
6. Computing.

In the technical project is where the quality indicators are established. Such indicators measure, quantitatively or qualitatively some of the quality components in official statistics: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence, completeness, cost and burden.

The Register of technical projects and questionnaires uses an application that files away the operation technical documentation and generates all the administrative documentation on the inscription of projects therein. The application has several user profiles: Register administrator, technician and Manager. The technical user reviews a project presented and issues a technical report. The Manager user, once he/she has read the report, decides on whether to proceed to the approval and inscription in the Register.

The filing of the documentation is carried out within a Sharepoint environment to which access is granted to the members of the whole of the Basque Statistical Organisation who are the producers of official statistics. Documentation modifications are to be carried out by means of the Register application. The Sharepoint site is also a forum for all the official statistics producers, with all the available methodological information and best methods manuals.

marina_ayestaran@eustat.es, cristina_prado@eustat.es

Subsystem QUALITY in Statistical Metainformation System

Jitka Prokop - CZSO, Czech Republic

Subsystem QUALITY (Q subsystem) is a part of newly projected statistical metainformation system (SMS) of the Czech Statistical Office (CZSO).

The aim is to create methodology and tools for timely monitoring of information on quality of statistical process (i.e. a regular statistical activity, from proposal to dissemination), its inputs and outputs, for the purposes of quality reports, EFQM reports and other feedback for management on all levels, dissemination and documentation. Q subsystem shall help harmonize quality assessment (used indicators and measures, monitored metainformation), support activities for improvement of statistical quality, and enable comfort comparisons.

Q subsystem is in compliance with strategy of the CZSO. It is based on the quality concept of the European statistical system (ESS), structured by the ESS quality dimensions. Themes cover all phases of statistical process.

The core function of Q subsystem is to enable views from data warehouse (DWH). They monitor three levels: statistical process, particular data processing (one run), and statistical variables. General structure is the following: theme – section – (row – column) – item. By item we mean measure, indicator or other metadata on quality.

Structure of the views containing all items that can be used is called General QF map (QF – quality formular). For a particular type of statistical process, Specific QF map might be designed and derived from the General QF map.

A QF map, filled with concrete data and metadata, is called the Data QF (the DQF). Each run ending by releasing of data out of the process has its own version of the DQF. If a figure in an item is valid for more DQFs within a process, there is only one record in database. Historical records remain in database.

Figures enter Q subsystem in parts, as soon as they are available. During process they might be updated. Values are filled manually or, for as many items as possible, there is an automatic link to another SMS subsystem in DWH (mainly PROCESSES), either direct one or with calculation formula.

Software for calculation of key quality indicators for accuracy (i.e. response rate, coefficient of variation, imputation rate, size of revisions), applicable for most of statistical surveys, and for EFQM performance indicators on data quality has been prepared and is used in Q subsystem.

Codings (usually in Czech and English) are stored in the SMS subsystems VARIABLES and CLASSIFICATIONS. They cover e.g. quality indicators and measures, revisions, survey methodologies, data protection, products/publications, users and their categories etc.

Different roles within Q subsystem are defined. Methodologist as an administrator defines new content of QF map in cooperation with statisticians. Survey manager is responsible for filled figures, he can ask methodologist for cooperations. Direct access to Q subsystem is available only within the CZSO; internal users of Q subsystem are divided into several groups depending what outputs they need. For future, we consider use of Q subsystem as a source for providing information externally.

QUALITY METADATA + REPORTING

MetaPlus a Metadata Tool for the Production Process

Klas Blomqvist - Statistics Sweden

MetaPlus is a tool primarily developed for documenting micro data. The data model is compliant with ISO 11179, which makes it possible to develop the system further to handle all metadata aspects of the statistical life cycle. The paper focus-

METADATA +
HARMONIZATION

es on the experiences so far from working with MetaPlus, in particular registering metadata. Since the subject matter departments are responsible for registering the contents of the documentations in MetaPlus the paper stresses the need for building a support organisation with the KMI (classifications, metadata and content harmonisation) group at the Process department and the documentation network. Connections to other Metadata systems, such as the archiving functionality and future plans are also mentioned.

Variables System – the Bridge between Metadata and Dissemination

Teodora Monica Isfan - Statistics Portugal

“Metadata is the key to ensure that information will survive and continue to be accessible into the future.”

The need for metadata standards, particularly definitions that are consistent among the myriad of subject matters of a statistical agency is more urgent today than it has ever been. More recently, the availability of diverse types of data and information on the Internet means that it is possible for users to view and compare data from one subject matter area with that of another. The development of a dissemination database to support the Portal of Statistics of Portugal, forced the Metadata Unit to the registry of metadata of statistical indicators directly in the variables database, so users can access not only data but also the description of their meaning: definition, calculating formula, unit of measure, population studied, and the type of data they represent, among others. The main goal of this paper is to present the principles used to guide the design, necessary standards, the tools and the solutions to realise through an automatic way the association metadata – dissemination database, without any lose of quality.

monica.isfan@ine.pt

METADATA + VARIABLES +
STATISTICAL INDICATORS

Associated papers

Development of Metadata System in CSO (Hungary)

Zsófia Ercsey - Central Statistical Office, Hungary

Aim: to create an integrated metadata system and make it available on internet for data users.

How: launching 3 projects:

1. Concepts, definitions, nomenclatures
2. Methodological documentation on statistics
3. Metadata on website of CSO.

Results:

1. Concepts, definitions have been described and loaded in meta database.
2. Nomenclatures have been standardized and loaded in meta database
3. Methodological documentations on statistics have been compiled by a standard scheme and loaded in meta database
4. Design of website of CSO.