INSTITUTO NACIONAL DE ESTADISTICA

Standard for documenting production processes of statistical operations at INE

Standardised reports of metadata processes

7 April 2015

1 Justification for the need of the standard

The INE considers necessary to strengthen, structure and systematise the effort it makes in documenting the production systems it uses for the production of statistical operations.

This aspect is essential in any institution aimed at production, since it provides numerous guarantees to the Institution. Worth mentioning among others are that:

– It systematises the knowledge of the way in which it is produced in the Institution, this way guaranteeing that the process is executed in the way in which it is determined.

– It provides guarantees of continuity over time facing the mobility of persons.

– It allows the analysis, assessment and improvement of the production processes.

At the same time, it is the way to study where it is necessary or convenient to introduce standardised processes for specific phases, sub-processes or tasks of the process and the introduction of practices that make production more efficient. This need to progress towards a standardised and industrialised production process is specially urgent in a scope in which resources decrease and the need for information increases.

In order to carry out this strategy, it is convenient to follow the standards that allow comparing the way in which these statistical operations are produced within the INE, but also other producers of the Spanish Statistical System and the European Statistical System, always pursuing the improvement of efficiency.

In this context, GSBPM (Generic Statistical Business Process Model) is a consolidated international standard which is adopted by numerous statistical offices and international institutions, which proposes a process and sub-process structure of the statistical production model.

For all of these reasons, the INE has the institutional commitment, supported by previous agreements of the Board of Directors and demarcated in the improvement actions derived from Peer Review 2014-2015, to compile a metadata process standard, which based on GSBPM adapts to the objectives, needs and the production method of the INE, and achieving the implementation of that standard in all the statistical operations of the INE before July 2016.

2 Works developed for the compilation of the standard

The starting point for preparing the standard was a pilot experience to compile the documents of seven statistical operations following GSBPM that could not be fulfilled for the following reasons:

− There were three tasks that were not documented in all the operations.

− The level of detail varied a lot among operations and even among tasks within the same operation.

− There were tasks that were located in different sub-processes of the GSBPM depending on each operation, this way complicating their analysis.

To resolve these limitations and proposing a documenting standard for the INE, in November 2014 the Board of Directors created a work group formed by representatives of the following units: methodology (which coordinated the group), sampling, data collection, IT, dissemination, quality, municipal register, price statistics and household budgets and population statistics. This group was ordered to compile a process metadata file that would imply the adaptation of GSBPM to the needs and the production method of the INE.

The group suggested that in order to correct the problems detected, it was necessary to create a third level of tasks that would complement the GSBPM. This decision was not made with the objective of increasing the documents to be produced, but with the purpose of guaranteeing the completeness of the task list that forms each sub-process of GSBPM and that their description is sufficient and similar among operations. In this way, the documents of all operations will have an equal structure and a homogeneous level of detail, allowing for the analysis of this information so as to assess and improve the processes in the search for a production system that is as integrated and standardised as possible.

Moreover, in order to view and understand the production process, the group included in the standard the configuration of work flows at task level following BPMN 2.0 (Business Process Modelling Notation).

As a collection tool, the group agreed upon the updating of the reference metadata editor, making it the only entry point for metadata when incorporating the new form for the completion of the process metadata. It was also agreed that the work flow be designed using the Bizagi Modeler tool, which is free and user-friendly software.

The proposal of the group was sent to all units and subdirectorates related with statistical production so as to collect their opinion on it. Their comments are compiled in the final proposal of the standard, which is the following.

3 Proposal for a standard

1. The documents of the production processes of all statistical operations of[1] the INE will be developed following the process metadata structure that appears in annex I and that consists in a third level breakdown of the GSBPM.

2. The following elements shall be specified for each task: input, output, process, tool, documents and responsible unit.

3. The task description will be completed with the work flow configuration at task level.

4. For the completion of the files, there is a computer application (metadata editor) and support documents.

---

[1] As collected in the Inventory of Statistical Operations

4 Procedure for the revision of the standard

The revision and updating of this standard must be approved by the Board of Directors of the INE.

Annex I

Structure of the standardised report of process metadata (task list)

Task list of GSBPM phase 1

1. Specify need

1.1 Identify needs

1.1.1 Identify external needs for information
1.1.2 Identify internal needs for information
1.1.3 Compare similar statistical operations
1.1.4 Identify restrictions
1.1.99 Other tasks

1.2 Consult & confirm needs

1.2.1 Contact users
1.2.99 Other tasks

1.3 Establish output objectives

1.3.1 Identify population aggregates
1.3.2 Identify dissemination domains
1.3.3 Identify dissemination metadata (for dissemination)
1.3.4 Identify dissemination time scope
1.3.5 Identify requirements of statistical disclosure control
1.3.6 Identify requirements for seasonal /calendar-effect adjustment
1.3.99 Other tasks

1.4 Identify concepts

1.4.1 Identify population/statistical units
1.4.2 Identify variables
1.4.3 Identify directories
1.4.4 Identify administrative registers
1.4.99 Other tasks

1.5 Identify concepts

1.5.1 Analyse directories
1.5.2 Analyse administrative registers
1.5.3 Check availability of other sources
1.5.4 Analyse other statistical operations
1.5.99 Other tasks

1.6 Identify concepts

1.6.1 Prepare feasibility report
1.6.2 Prepare statistical project
1.6.3 Prepare planning and budget
1.6.4 Communicate new projects to users
1.6.99 Other tasks

Task list of GSBPM phase 2

2. Design

2.1 Design outputs s

2.1.1 Design estimate tabulations
2.1.2 Design metadata for dissemination
2.1.3 Design dissemination time scope
2.1.4 Design dissemination formats
2.1.99 Other tasks

2.2 Design variable descriptions

2.2.1 Operationalise statistical units
2.2.2 Operationalise variables
2.2.3 Operationalise population aggregates
2.2.4 Define derived variable
2.2.5 Design statistical treatment of administrative registers
2.2.6 Design statistical treatment of other sources
2.2.99 Other tasks

2.3 Design collection

2.3.1 Design annual planning for collection
2.3.2 Select collection mode
2.3.3 Design questionnaire
2.3.4 Design administrative register schema
2.3.5 Design schema for other collection modes
2.3.6 Design collection instruments
2.3.7 Design material supporting collection
2.3.8 Design paradata
2.3.9 Design data collection management application
2.3.10 Design linguistic aspects of collection
2.3.11 Design H.R allocation for collection
2.3.12 Design H.R selection/training plan
2.3.13 Determine collection infrastructure
2.3.14 Design collection-related contracts
2.3.15 Design collaboration agreement
2.3.16 Design promotion action for collection
2.3.99 Other tasks

2.4 Design frame & sample

2.4.1 Design /update population frame
2.4.2 Identify/update sampling method
2.4.3 Determine sample size and sample allocation
2.4.4 Determine/update sample selection and sample coordination
2.4.5 Design/update sample extraction
2.4.99 Other tasks

2.5 Design processing & analysis

2.5.1 Design coding
2.5.2 Design error detection
2.5.3 Design error treatment
2.5.4 Design validation of edited sample
2.5.5 Design integration of files from the same source
2.5.6 Design integration of diverse sources
2.5.7 Design index/aggregate weights calculation
2.5.8 Design population aggregates estimators
2.5.9 Design corrected population aggregates estimators
2.5.10 Design statistical treatment of administrative registers
2.5.11 Design statistical treatment of other sources
2.5.12 Design variance estimators
2.5.13 Design time series updating
2.5.14 Design seasonal/calendar-effects adjustment
2.5.15 Design statistical disclosure control
2.5.99 Other tasks

2.6 Design production systems & work-flows

2.6.1 Design work-flows
2.6.2 Prepare work schedule
2.6.3 Design process coordination
2.6.4 Design data and metadata databases
2.6.5 Design microdata backups
2.6.6 Design other information system components
2.6.99 Other tasks

Task list of GSBPM phase 3

3 Build

3.1 Build collection instrument

3.1.1 Prepare questionnaire
3.1.2 Build collection instrument
3.1.3 Build collection instrument for administrative registers
3.1.4 Build collection instrument for other sources
3.1.5 Prepare material supporting collection
3.1.6 Build/configure collection infrastructure
3.1.7 Program collection metadata
3.1.8 Build collection management applications
3.1.9 Include linguistic aspects of collection
3.1.10 Prepare material for H.R training plan
3.1.11 Prepare collection contracts
3.1.12 Prepare collaboration agreements
3.1.13 Build tools for promotion actions for collection
3.1.99 Other tasks

3.2 Build or enhance process components

3.2.1 Build IS components for the construction of the frame population
3.2.2 Build IS components for the sampling design
3.2.3 Program sample extraction
3.2.4 Build IS components for coding
3.2.5 Build IS components for error detection
3.2.6 Build IS components for error treatment
3.2.7 Build IS components for the validation of the edited sample
3.2.8 Build IS components for the integration of files from the same source
3.2.9 Build IS components for the integration of diverse sources
3.2.10 Build IS components for the calculation of aggregate/index weights factors
3.2.11 Program estimators of population aggregates
3.2.12 Program corrections of estimators of population aggregates
3.2.13 Program variance estimators
3.2.14 Build IS components for the tabulations of estimates
3.2.15 Build IS components for the calculation of derived variables
3.2.16 Build IS components for updating time series
3.2.17 Build IS components for seasonal/calendar-effect adjustment
3.2.18 Build IS components for statistical disclosure control
3.2.19 Build IS components for microdata and metadata databases
3.2.20 Build IS components for microdata backup
3.2.99 Other tasks

3.3 Build or enhance dissemination components

3.3.1 Build system for the presentation of outputs
3.3.2 Build IS components for dissemination data and metadata databases

3.3.3 Build other dissemination tools
3.3.99 Other tasks

3.4 Configure work-flows

3.4.1 Configure data collection work-flows
3.4.2 Configure data processing work-flows
3.4.3 Configure dissemination work-flows
3.4.99 Other tasks

3.5 Test production system

3.5.1 Test tools for sample selection and construction of the population framework
3.5.2 Test data collection tools
3.5.3 Test processing tools
3.5.4 Test dissemination tools
3.5.99 Other tasks

3.6 Test statistical business process

3.6.1 Execute pilot test
3.6.2 Test jointly IS components
3.6.99 Other tasks

3.7 Finalise production system

3.7.1 Execute H.R. training
3.7.2 Execute maintenance
3.7.99 Other tasks

Task list of GSBPM phase 4

4 Collect[1]

4.1 Create frame & select sample

4.1.1 Execute construction of frame population
4.1.2 Execute sample extraction
4.1.99 Other tasks

4.2 Set up collection

4.2.1 Allocate collection material
4.2.2 Allocate/install collection infrastructure
4.2.3 Initialise collection management applications
4.2.4 Allocate H.R for collection
4.2.5 Execute H.R training plan for collection
4.2.6 Execute promotional action for collection
4.2.7 Guarantee hiring conditions for collection
4.2.8 Request administrative information
4.2.99 Other tasks

4.3 Run collection

4.3.1 Initialise collection
4.3.2 Attend to questions on collection
4.3.3 Execute data entry
4.3.4 Execute collection claims
4.3.5 Execute collection monitoring
4.3.6 Execute collection inspection
4.3.7 Execute collection of administrative registers
4.3.8 Execute collection from other sources
4.3.99 Other tasks

4.4 Finalise collection

4.4.1 Prepare data collection report
4.4.2 Prepare report on the attainment of administrative registers
4.4.3 Prepare report on the attainment from other sources
4.4.4 Transmit collected data and metadata
4.4.5 Execute archiving of collection documents
4.4.99 Other tasks

---

[1] In the Spanish version an explicit distinction between collection of survey data (*recoger*) and collection of administrative registers (*obtener*) is made; in this English version we have respected the original overall term *collection*.

Task list of GSBPM phase 5

5 Process

5.1 Integrate data

5.1.1 Execute integration of data coming from the same collection mode
5.1.2 Execute integration of data coming from diverse collection modes
5.1.99 Other task

5.2 Classify & code

5.2.1 Execute code
5.2.2 Execute coding quality control
5.2.99 Other tasks

5.3-4 Review & validate. Edit & impute

5.3.1 Execute error detection and treatment (input)
5.3.2 Execute error detection and treatment (output)
5.3.3 Prepare edit and imputation report
5.3.99 Other tasks

5.5 Deriving new variables and units

5.5.1 Compute new statistical units
5.5.2 Compute derived variables
5.5.99 Other tasks

5.6 Calculate weights

5.6.1 Calculate index/aggregate weights
5.6.99 Other task

5.7 Calculate aggregates

5.7.1 Calculate (corrected) estimators of population aggregates
5.7.2 Calculate variance estimators and coefficients of variation
5.7.3 Execute quality control over estimator calculations
5.7.99 Other tasks

5.8 Finalise data files

5.8.1 Update microdata and aggregate databases
5.8.2 Execute microdata backup
5.8.99 Other tasks

Task list of GSBPM phase 6

6 Analyse

6.1 Prepare draft outputs

6.1.1 Execute time series updating
6.1.2 Execute seasonal/calendar-effect adjustments
6.1.3 Prepare tabulations estimate
6.1.4 Prepare microdata files for users
6.1.5 Prepare metadata reports for dissemination
6.1.99 Other tasks

6.2-3 Validate results. Interpret and explain outputs

6.2.1 Execute error detection and treatment (macro)
6.2.2 Execute validation of edited sample
6.2.3 Interpret outputs
6.2.99 Other tasks

6.4 Apply disclosure control

6.4.1 Drop out direct identification variables
6.4.2 Assess identification risk
6.4.3 Apply information protection
6.4.4 Configure secure access centres
6.4.99 Other tasks

6.5 Finalise outputs

6.5.1 Prepare internal reports
6.5.99 Other tasks

Task list of GSBPM phase 7

7 Disseminate

7.1 Update output systems

7.1.1 Execute validation of dissemination products
7.1.2 Update dissemination databases
7.1.3 Guarantee data-metadata link
7.1.99 Other tasks

7.2 Produce dissemination products

7.2.1 Apply dissemination format
7.2.2 Execute information compilation
7.2.3 Prepare element of dissemination product
7.2.4 Validate element of dissemination product
7.2.5 Prepare dissemination product
7.2.6 Approve dissemination product
7.2.7 Prepare press release
7.2.99 Other tasks

7.3 Manage release of dissemination products

7.3.1 Manage dissemination calendar
7.3.2 Transmit dissemination products
7.3.3 Publish dissemination products
7.3.4 Manage errors in dissemination products
7.3.5 Distribute customised dissemination products
7.3.6 Publish metadata for dissemination
7.3.99 Other tasks

7.4 Promote dissemination products

7.4.1 Manage activity on social networks
7.4.2 Manage informative material
7.4.3 Execute promotional action on dissemination
7.4.4 Maintain dissemination channels
7.4.5 Manage editorial programme
7.4.99 Other tasks

7.5 Manage user support

7.5.1 Manage customised dissemination product
7.5.99 Other tasks

Task list of GSBPM phase 8

8 Evaluate

8.1 Gather evaluation inputs

8.1.1Gather reports
8.1.2 Gather metadata for quality evaluation
8.1.3 Gather suggestions from H.R.
8.1.99 Other tasks

8.2 Conduct evaluation

8.2.1 Identify process errors
8.2.2 Compare quality indicators
8.2.99 Other tasks

8.3 Agree an action plan

8.3.1 Prepare action plan for improvement
8.3.2 Prepare documentation about the action plan for improvement
8.3.3 Execute follow-up of the action plan for improvement
8.3.99 Other tasks