

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN
STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD) STATISTICS DIRECTORATE

Workshop on Statistical Metadata (METIS)

(Geneva, Switzerland 5-7 October 2011)

Session 4: Topic IV: The GSBPM and Process Quality Management

The business process models and quality issues at the Hungarian Central Statistical Office (HCSO)

Mr. Csaba Ábry, Hungarian Central Statistical Office

I. Introduction

1. The aim of this contribution paper is to show, how the Hungarian Central Statistical Office uses its business process models and GSBPM for its statistical data production (especially for quality issues) and for the management of this activity.

The paper is broken down to four main chapters:

I.1. Present situation at the HCSO

It gives some historical information on the work done so far in this issue, including the mapping procedure of the domestic business process models and the GSBPM.

I.2. Quality of metadata

This chapter provides general information on the metainformation system of the HCSO and the quality of the metadata in this system.

I.3. Quality of statistical data

This chapter gives information on the used quality indicators (both product quality indicators, and process quality indicators) used and their connection with the business process models used at the HCSO.

I.4. Future works

This chapter summarises the future work that has to be done by the HCSO.

II. Present situation at the HCSO

2. At the Hungarian Central Statistical Office there are several business process models according to the different issues (quality, process of statistical data, IT business model, work plan, resource planning, working time register, etc). These process models were determined in the framework of different projects in different times, e.g. the IT business model, done by our IT Department was worked out decades ago, since the introduction of our metainformation system goes back to the end of the 70's.

3. The above mentioned projects aimed different issues e.g. elaborating of quality guidelines, quality indicators, or the establishment of the Uniform Data Process System. During these

projects the elaborated process models were always mapped to each other and several stakeholders had the opportunity to comment on those process models before establishing them. But not all of the comments were taken into account.

4. The basic aim of the mappings done so far is to have only one business process at the office (if it is possible). At our office the IT business model seems to be the most likely candidate for being a standard process model, since it covers all phases of the statistical process. But there is a need for further analysis for that. In 2010 a draft mapping was compiled and discussed by several stakeholders of statistics in the HCSO.

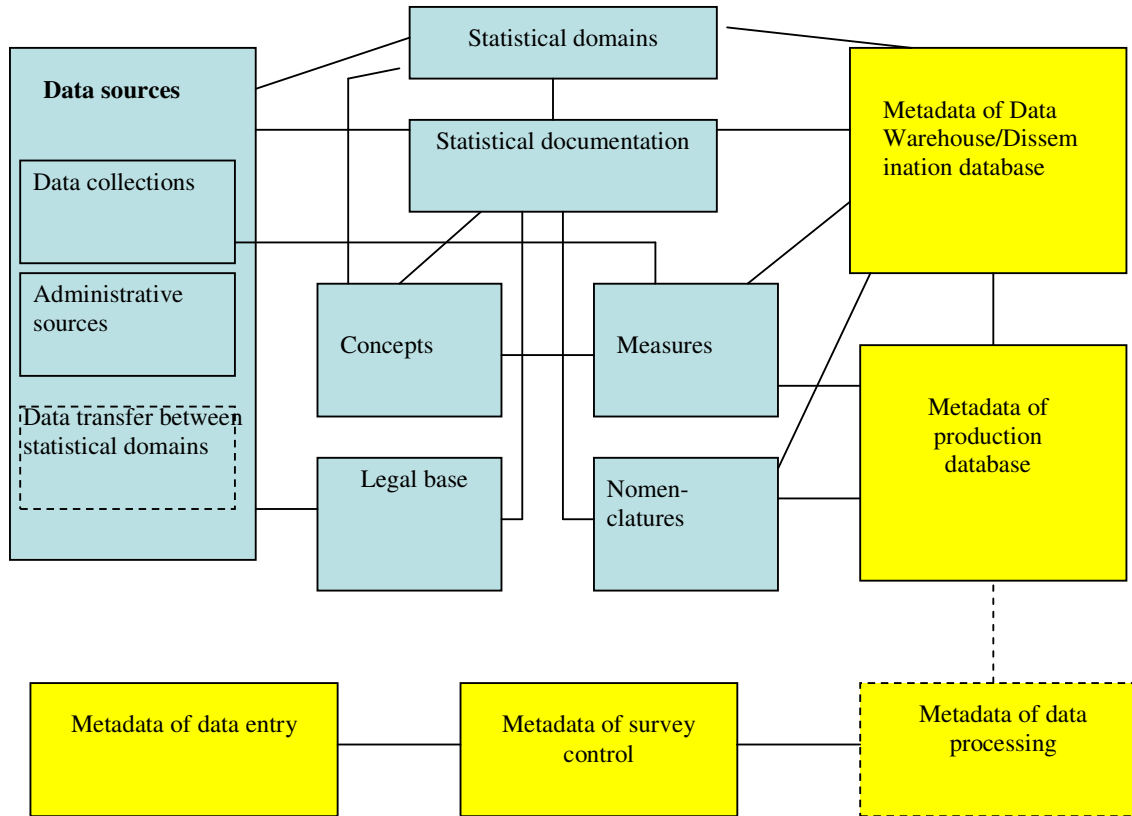
5. As the GSBPM has to be taken into consideration a new draft mapping was compiled this year (see Annex 1.). The Annex 1 only shows the mapping of process phases concerning quality to the IT business model and GSBPM. The GSBPM has been examined in a detailed way. As a result, it can be highlighted that it cannot be mapped one-by-one to the business process models of the HCSO, therefore currently it is used only as a 'vocabulary' in order to speak a common language at international level. Since this contribution paper is made for METIS meeting, the process phases of GSBPM are used.

III. Quality of metadata

6. In order to understand the metainformation system of the HCSO some general information on this system is needed. This system is integrated; it covers almost all of the processes of the HCSO. Only the activities of management of the HCSO (e.g. bookkeeping) are excluded. Metadata can only be used from this metainformation system (measures, classifications, nomenclatures, concepts, etc).

7. Unfortunately there is not one formal section or department, which has the responsibility for managing all of the metadata established at the HCSO. The metainformation system of the HCSO therefore can be called decentralised. As I mentioned in the previous chapter, the responsibility is shared by sub-systems of the metainformation system concerning IT tasks. Within these sub-systems the detailed tasks are shared by the IT business process model. See figure below (the yellow boxes cover process metadata as well):

Figure 1: Main subsystems of metainformation system



8. Since the responsibility is shared, quality aspects on metadata can be found in different IT documents. The reason is that those aspects have been worked out by different sections of the IT department of the HCSO according to their responsibility. What kind of quality aspects do those different documents cover? The IT documents on the quality aspects of metadata are broken down by GSBPM process phase.

Specify needs

9. There is a document on planning of surveys, which gives information on the planned surveys e.g. the reason for this survey, costs, etc.

Design

10. The Guideline for designing statistical database contains the stipulations on the establishment of metadata. It regulates the usage of metadata at database designing and the common rules for using them (e.g.: coding, correspondence tables, measures, nomenclatures,

concepts, etc). Metadata from the metainformation system of the HCSO have to be used for the whole business process.

Build

11. For the Build phase the most important document is the document on naming convention and the Guideline for designing statistical database. The document on naming convention stipulates the database themes of the HCSO. It means that the further metadata established for a certain theme can only coded according that certain database theme.

12. For example:

There is the database theme 'M', which is for Agriculture. The subtheme 'MA' is for animal husbandry. The classifications, measures, database tables have to use this code for the first two digit of the identifier (e.g. measure 'MAAA001' means Number of livestock and nomenclature 'MA01' means Species of livestock).

13. At the HCSO, there is a Quality Assurance Application for the improvement of data entry sub-system. This application is to support the work-flow of the improvement. It includes the responsibility issues, the deadlines, the results and the feedback for testing as well.

14. In this phase a correspondence between the logical and the physical description of a questionnaire has to be made. It is checked by the IT expert and loads into the metadatabase. This correspondence is used by our ADÉL system, which supports the data entry phase.

Collect

15. At the HCSO the IT applications ADÉL, and GÉSA support this process phase. In the IT documentation of those applications there are prescriptions for the commonly used metadata e.g.: type of errors, responsibility of the activities carried out during this process phase, etc. Those check points are built in the applications.

Process, Analyse

16. At the HCSO there is a Uniform Data Process System (EAR) under development, which has built in quality aspects for metadata. Since this system is also a metadata driven system, only metadata from the metainformation system of the HCSO can be used during the process phase. The system checks whether the subject-matter statistician uses the correct metadata (measures, classifications, etc.). There is a correspondence table in the system, which has to be used for mapping the defined database table and the metadata from the metadata system. For that phase the general rules for establishing metadata for aggregates are valid.

17. Further IT application for these phases is Demetra, which supports the activity of seasonal adjustment. There are built-in quality indicators, which helps the subject-matter statistician to decide whether the result got form this application is correct or not.

Disseminate

18. In the sub-system Data warehouse there are further check points built in the database. E.g: definitions of measures have to be filled, connected concepts and their definitions have to be filled also. The content is checked by the Methodological Department, since the control of contents cannot be automated. There is a planning template for the data cubes, which contains

all of the metadata on the certain data cube (e.g.: denomination, used measures, used classifications, etc). Those metadata have to be put into the metadatabase, according to the naming convention of the HCSO and they have to be connected with each other.

19. Concerning the sub-systems ‘Concepts’ and ‘Nomenclatures (classifications)’, there are built in controls in the database. If a subject-matter statistician wants to make a concept, a classification or methodological description of subject-matter statistics, and registers visible on the website of the HCSO, then it is checked whether the English denomination the English descriptions are loaded into the database. Before those items appear on the website there is a lecture process before. It aims both the content and the language control of those. In case of concepts, classifications, it is also checked whether a concept or a classification is with the same denomination and/or same content is already available in the database.

Archive

20. There is no document on this process phase aiming the quality of metadata of archiving.

Evaluate

21. There is no document on this process phase aiming the quality of metadata of evaluation.

IV. Business process model and process quality at the HCSO

4.1. Process quality

22. There are several standards at HCSO covering process quality. Quality guidelines were developed as quality requirements, process variables were identified for different process stages to monitor quality and an assessment tool prepared. First the elements of the statistical value chain were defined, based on this a production process model was created. The handbook entitled Quality Guidelines covering all these production process steps was compiled. The handbook provides some brief guidelines on the principles, recommendations that should be kept in mind to achieve the best quality during these production process steps.

23. The structure of the handbook follows the statistical production process therefore it consists of the following chapters:

- Register
- Frame
- Objectives, Uses and Users of the survey
- Concepts, Definitions and Classifications
- Accessing administrative data
- Sampling design
- Questionnaire design
- Data collection
- Data capture, micro validation, editing
- Imputation
- Weighting, estimation, sampling error computation
- Index number construction
- Macro validation
- Seasonal adjustment
- Further analysis
- Confidentiality and disclosure

- Dissemination
- Data archiving
- Assessment, Review and Feedback

24. The structure of the different chapters is the same: definitions, principles, quality guidelines. An annex contains the bibliography of actual references related to the production process steps.

4.2. Process quality indicators

25. Process quality indicators (PQI) provide information on the quality of certain production process steps so that assigned people could intervene during the process. The process steps in the manual for the elaboration of indicators are in line with production process model of HCSO. The manual for PQIs contains guidelines for the elaboration of specified process quality indicators, since the data-collections can be considerably different. A set of process quality indicators was developed according to the phases of business process model determined at the handbook on quality guidelines. IT systems used at HCSO support calculation of several quality indicators, new IT developments take into consideration their calculation. In the long term a common database should be created to store quality indicators.

26. In the following some indicators are listed by GSBPM process phases.

27 Specify needs

Share of not fulfilled users' demand
Number of methods used for collecting user's needs

For the 'Design' and 'Build' phase the supporting IT application is GÉSA The application GÉSA is for supporting the sub-system 'Survey control' of the HCSO's metainformation system.

28. Design (supporting IT application: GÉSA)

Under/overcoverage of a register
Accuracy of register data
Errors of sample frames
Number of methods used for questionnaire testing, etc.

29. Build (supporting IT application: GÉSA)

Differences between domestic and international standard concepts
Comparability of administrative data sources
Completeness of administrative data sources
Average time for filling a questionnaire
Data suppliers engaged in testing a data collection
Completeness of instructions, etc.

30. Collect (supporting IT application: ADÉL, LAKOS, GÉSA)

The above mentioned IT application, ADÉL is for the data entry sub-system of the metainformation system of the HCSO. This is for institutional data collections. For population data collections a new application is under development, the so called LAKOS.

Punctuality of sampling
Average number of questionnaires by surveyors
Rate of response/ non-response
Rate of controlled interviews
Rate of questionnaires collected electronically

31. Process (supporting IT application EAR)

EAR is the IT application under development for data processing. Now there are several production databases for data processing. EAR is developed for having a uniform data production database.

Number of errors at data entry
Rate of miscoded records
Share of outliers by collected data
Rate of imputation (item and unit)

32. Analyse (supporting IT application EAR)

EAR is the IT application under development for data processing. Now there are several production databases for data processing. EAR is developed for having a uniform data production database.

Coefficient of variation
Design effect
Number of methods used for validation
Risks at disclosure control
Rate of revision
Share of outliers in the time series, etc.

33. Disseminate (supporting IT application Data warehouse and Dissemination database. Other dissemination form, from which data can be used for calculating these indicators: Stadat (ready-made tables, STAR databases)

Timeliness (preliminary and final data)
Number of downloads
Rate of missed methodological descriptions,
Average size of revisions, etc

34. Archive

Share of archived documents
Share of archived data sets

35. Evaluate

There are no indicators for that process phase. But there is an application aiming the feedback of errors for processes or for products. This is the so called TETRISZ IT application, which gives the opportunity to the subject-matter statisticians to provide information on the quality of a certain process phase, or rather product by using the certain identifier of a product.

Quality of statistical products can be classed under evaluation phase. A set of standard quality indicators were elaborated, template of standards quality report was developed including quality indicators (see: Annex 2.) concerning the products as statistical data and analyses. Other developments have been carried out for improvement similar tools for some special products (administrative data, registers).

36. Overarching indicators for the processes (valid for each process):

Timeliness of a certain activity
Presence of resources

Having the quality requirement and measurement tools for the statistical process, the next step is the need for quality assessment tools.

A standard for self assessment has been developed based on DESAP checklist, adopted according to HCSO quality guidelines.

V. Possible future works

37. As it can be seen from this document there are still some works to be done:

- Further analysis of the existing business process models and their mapping is needed, in order to have only one business process model, with regard to GSBPM.
- Further work might be to determine all of the products, which are inputs and outputs of a process phase. It is needed for further improvement of product quality indicators. This might also lead to a more precise description of a process phase.
- Another work is to load the quality indicators to the database and to automate their calculation, which also means an IT development at the HCSO.
- The quality reports and indicators are advisable to make visible on the website of the HCSO, but further analysis is needed in that case as well.

VI. Sources:

38. Sources include

- Generic Statistical Business Process Model METIS Workshop on the Statistical Business Process and Case Studies (Lisbon, Portugal, 11-13 March 2009)
- Catalogue of process quality indicators, HCSO, 2009
- Quality guideline at the HCSO, HCSO, 2007
- Product quality indicators, HCSO, 2008
- Naming convention of the HCSO metainformation system, HCSO, 2006
- Documentation on ADÉL (data entry) sub-system, HCSO, 2001
- Guideline on uniform questionnaire design, HCSO, 2006
- Quality assurance of questionnaire design, HCSO, KSH-School training paper, 2008
- Quality assurance for centralised sending of questionnaires, HCSO, 2005
- Controls in the ADÉL (data entry) system, HCSO, 2001
- Documentation of Quality Assurance Application for database improvements, HCSO, 2010
- Project on the establishment of the Uniform Data Process System (EAR), HCSO, 2009
- Documentation on TETRISZ (uniform communication tool) application, HCSO, 2010
- Documentation on the data warehouse of the HCSO, HCSO, 2009
- Guideline for designing statistical database, HCSO, 2006
- Project on integrating spatial data into registers, 2011

Annexes

Annex 1. Mapping of processes at the HCSO

IT business process model at the HCSO (management of metadata are included in every process phase, only maintenance of metadata are highlighted as separate phase)		Quality guideline at the HCSO (process phases important for quality aspects)		GSBPM	
Code.	Denomination	Steps	Denomination	GSBPM sub-processes	GSPPM process phases
1	Design	a)	Selecting Register;	2.4. Design frame & sample methodology (partly)	2. Design (partly)
		b)	Determining sample frame (partly)	2.4. Design frame & sample methodology (partly)	
		c)	Determining aims, users, stakeholders	1. Specify needs	1. Specify needs
		d)	Determining concepts, definitions, classifications (partly)	2.2 Design variable descriptions (partly)	2. Design (partly)
		e)	Taking stock of available administrative data sources	2.3. Design data collection methodology (partly)	
		f)	Designing frame and sample	2.4. Design frame & sample methodology	3. Build
		g)	Designing questionnaire	2.2 Design variable descriptions (partly); 3.1. Design data collection instrument	
				2.5. Design Data processing methodology	
				2.6. Design production systems flows	
				3.1. Build data collection instrument	
				3.2. Build enhance process components	
				3.3. Configure workflows	
				3.4. Test production system	
				3.5. Test statistical business process	
		3.6. Finalize production system			
2	Management of registers				

IT business process model at the HCSO (management of metadata are included in every process phase, only maintenance of metadata are highlighted as separate phase)		Quality guideline at the HCSO (process phases important for quality aspects)		GSBPM		
Code.	Denomination	Steps	Denomination	GSBPM sub-processes	GSPPM process phases	
3	Maintenance of metadata				Metadata management	
4	Survey control	h)	Survey control and data collection	2.4. Design frame & sample methodology	2. Design (partly)	
				4.1. Select sample;	4. Collect (partly)	
				4.2. Set up collection		
				4.3. Run collection		
				4.4. Finalize collection		
5	Data editing	i)	Data editing	4.4. Final collection	4. Collect (partly)	
				5.1. Integrate data	5. Process (partly)	
				5.2. Classify and code		
				5.3. Review, validate and edit		
6	Process	j)	Imputation	5.4. Impute	5. Process (partly)	
				5.5. Derive new variables and statistical unit		
			5.6. Calculate weights			
		k) Weighting, estimation, sampling error computation	5.7. Calculate aggregates			
			5.8. Finalize data files			
		k) Weighting, estimation, sampling error computation	5.6. Calculate weights			
		l)	Index number construction;		6.1. Prepare draft outputs	6. Analyse (partly)
m)	Macrovalidation;		6.2. Validate outputs			
7	Query from database	o)	Further analysis;	6.3. Scrutinize and explain	6. Analyse (partly)	
				6.5. Finalize outputs		
		p)	Confidentiality and disclosure control;	6.4 Apply disclosure control		
8	Dissemination	q)	Dissemination;	7.1. Update output systems	7. Disseminate	
				7.2. Produce dissemination products		
				7.3. Manage release of dissemination products		

IT business process model at the HCSO (management of metadata are included in every process phase, only maintenance of metadata are highlighted as separate phase)		Quality guideline at the HCSO (process phases important for quality aspects)		GSBPM	
Code.	Denomination	Steps	Denomination	GSBPM sub-processes	GSPPM process phases
				7.4. Promote dissemination products 7.5. Manage user support	
		r)	Data archiving;		8. Archive
		s)	Review, assessment, feedback		9. Evaluate

Annex 2. Quality components and the concerning general product quality indicators in the HCSO

<i>Quality aspect</i>	<i>General indicator</i>
Relevance	(R1) User satisfaction index
	(R2) Number of reference in the media
Accuracy	(P1) Coefficient of variation (CV)
	(P2) Unit response/non-response rates
	(P3) Imputation Shares/rates
	(P4) Edit failure Shares/rates
	(P5) Share of overcoverage and classification errors
	(P6) Average size of revisions
Timeliness	(I1) Time lag between the end of reference period and the date of the first/preliminary results
	(I2) Time lag between the end of reference period and the date of the final results
Punctuality	(IP1) Punctuality of publication
Accessibility	(H1) Types of tools for preliminary dissemination
	(H2) Number of database accesses
Clarity	(É1) Rate of completeness of metadata
Comparability and coherence	(Ö1) Length of comparable time-series