

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Luxembourg, 9-11 April 2008)

Topic 2 (iii) Metadata and the statistical cycle and Implementation

CASE STUDY: STATISTICS PORTUGAL

Submitted by Portugal¹

Revision History.....	2
Organization Details.....	2
1. INTRODUCTION.....	3
Metadata strategy	3
Current situation.....	5
2. STATISTICAL METADATA SYSTEMS AND THE STATISTICAL CYCLE.....	6
2.1 Statistical business process cycle	6
2.2 Current system(s)	10
2.3 Costs and Benefits.....	19
2.4 Implementation strategy	19
3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS.....	19
3.1 Metadata Classification	19
3.2 Metadata used/created at each phase.....	20
3.3 Metadata relevant to other business processes	22
4. SYSTEMS AND DESIGN ISSUES.....	22
4.1 IT Architecture	22
4.2 Metadata Management Tools	22
4.3 Standards and formats	23
4.4 Version control and revisions.....	23
4.5 Outsourcing versus in-house development	23
5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES	24
5.1 Overview of roles and responsibilities	24
5.2 Metadata management team.....	26
5.3 Training and knowledge management	27
5.4 Partnerships and cooperation	28
6. LESSONS LEARNED	29

¹ Prepared by Isabel Morgado (isabel.morgado@ine.pt) and Mónica Isfan (monica.isfan@ine.pt).

METIS COMMON METADATA FRAMEWORK (CMF)

PART C CASE STUDY

PORTUGAL / STATISTICS PORTUGAL

Revision History		
Date	Section(s) updated	Comment
18/02/2008		Version 1

Organization Details

Organization Name Statistics Portugal

Number of staff Statistics Portugal is comprised of a group of skilled professionals, employing 710 staff, (about 75% in Lisbon and 25% spread by four delegations across the mainland).

Contact person (for Metadata) Isabel Morgado
Head of Service/Systems and Metadata Service/Methodology and Information Systems Department
isabel.morgado@ine.pt
351 21 842 61 40

Organization structure

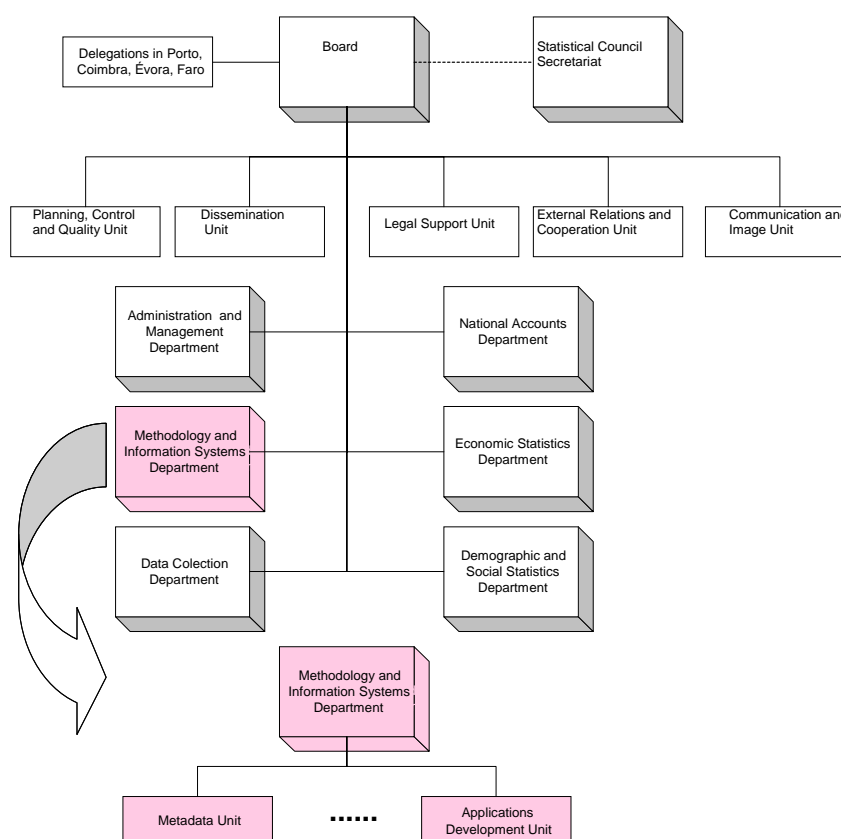


Fig.1. Organization structure

1. INTRODUCTION

Metadata strategy

The National Statistical System

The National Statistical System (NSS) consists of:

- The Statistical Council (SC);
- The National Statistical Institute - Statistics Portugal (SP).

The Statistical Council (SC) is the state body that supervises and coordinates the National Statistical System. Its duties include:

- “To guarantee the coordination of *the National Statistical System, approving the concepts, definitions, nomenclatures and other technical instruments of statistical coordination*” (Law 6/ 89 of 15 April – Diário da Republica 1st series no. 88).

This duty is carried out by the “Planning, Coordination and Dissemination” Standing Section (PCDSS), which has the power:

- “...to analyse and approve concepts, definitions, nomenclatures and other technical instruments of statistical coordination of the National Statistical System and to approve regular changes to these documents resulting from work done at the EU or national level”(Structure and Functioning of Statistical Council – SC Deliberation No. 286; 2005).

The job of the Statistics Portugal is to record, refine, coordinate and disseminate official data while taking into account the general guidelines laid down by the Statistical Council. It may also delegate these duties to other public departments, called delegated bodies.

SP has the responsibility to conceive and manage the statistical metadata system of the NSS, having as presumption that the concepts, classifications and other technical instruments of statistical coordination have to be approved by the SC. The metadata unit coordinates all the work related to the statistical metadata system.

Approval of concepts, classifications and methodological documentation

In these processes exists a strong interaction between the SP and the Statistical Council. SP compels all the information and prepares the documentation that sends to the SC for approval.

The SP centralises the statistical concepts used in its own and the delegated bodies’ statistical surveys in a database. These concepts are classified by subject areas and are entered into database with the status of “proposed concept”, when they are used for the first time. Groups of new concepts or changes to approved concepts are sent to the SC periodically for analysis and new approval. The SC has working groups by subject area to analyse them and recommend their approval to the PCDSS. After the approval, their status in the database is changed to “SC-approved concept” and, is of obligatory use whenever applicable.

The classifications used in all statistical activity, such as the Portuguese Classification of Economic Activities, National Classification of Occupations, National Classification of Goods and Services, Administrative Division Code and List of Countries are also approved by the SC for mandatory use in the NSS.

In 2005, the SP submitted to appreciation to the PCDSS a standard format for the methodological document for the NSS’s statistical surveys because it was considered to be a coordination instrument. The format was approved and adopted as mandatory in the NSS.

By December 2007, 75% of the surveys in the NSS were documented in this format.

Technical approval of surveys

The process for technical approval of surveys, which it is implemented at the SP's level without intervention of the SC, is closely linked to their life cycle and consists of the following stages:

- The preliminary methodological document and the questionnaire(s) produced in the methodological study are sent to the units directly involved or users of the results of surveys, the Planning Unit, the Data Collection Department and the Methodology and Information Systems Department for their opinion.
- At this point in the circuit, the Metadata Unit analyses the correct use of concepts and classifications approved by the SC, ensures correct application of the standard format in the methodological document also approved by the SC, analyses questionnaires, introduces new concepts into the concept base and issues an opinion on the basis of its analysis.
- The department responsible for the survey updates the methodological document and/or the questionnaire(s) with the proposed changes or justifies its rejection to the unit that proposed them, submitting then the new version of the methodological document and questionnaire(s) for approval by the Board.
- The Metadata Unit prepares a memo, on the basis of all the opinions of the different units and respective answers, to send to the Board, proposing their approval or rejection.
- The Board then approves or rejects the survey: if it approves them, the methodological document and the questionnaire(s) become final; if it rejects them, the process starts again.
- In the approval case, the Metadata Unit records the questionnaire(s) in the data collection instruments database, giving them a registration number and publishing the methodological documentation the Intranet and in the Official Statistics website.

Metadata dissemination – Statistics website

Concepts, classifications and methodological documents are available directly on the home page of the Statistics website.

The variables involved and the associated metadata must be recorded, previously, in the variables subsystem (one of the components of the Statistical Metadata System) so that the data on statistical indicators can be made available on the website.

Strategy for the metadata system

Organisations have attributed increasing importance to managing knowledge, as demonstrated by the growing implementation of metadata systems that systematise, standardise and formalise this knowledge so that it can be published within or outside the organisation.

In statistics organisations in particular, the systematisation of metadata is important because can be published along with statistical data so that they can be better understood, because gives to statisticians a clearer idea of the surveys that produce these data and for which they are responsible and, also plays a fundamental role in statistical coordination.

As a result, the aim of metadata systems in statistical organisations must be to support the entire life cycle of surveys and data and, makes sense to talk of a metadata life cycle.

In May 2002, the Metadata Unit submitted a document to the SP Board and Council of Directors laying down the general guidelines governing the NSS's Statistical Metadata System. Both bodies approved the document, which proposed the implementation of a system that:

- Will support surveys from their design to the dissemination of results;
- Will capture metadata for the system, from its origin, only once and with the possibility of being reused in other contexts;
- Consists of subsystems and components, so that they can be implemented in stages, with the possibility of navigating between the different components;
- Has decentralised management by the survey managers, with centralised coordination by the Metadata Unit;
- Allows different national and international bodies to exchange metadata;
- Supports other languages, such as English, in addition to Portuguese.

The implementation of this strategy has been included in medium-term work plans drawn up for the institution. The “General Guidelines on National Statistical Activity and Priorities for 2003-2007” defined as top priority:

- *“Implementing an integrated statistical metadata system — In the development of the statistical metadata system organised and coordinated by the SP, it is particularly important to design and implement an integrated classification management system, an integrated concepts management system and an integrated methodological document management system, define a model for a statistical subsystem and create support instruments for their implementation.”*
- *“Promoting the use of the statistical metadata system in the NSS.”*
- *“Improving user access to statistics - ...adapting the metadata system as a tool for accessing available information and making it easy to read and understand...”*

Two courses of action in the “General Guidelines on National Statistical Activity for 2008-2012” are devoted to the metadata system:

- *“To align the statistical metadata system with best international practices.”*
- *“To render the statistical metadata system appropriate to the needs of the interchange of metadata within the National Statistical System and the European Statistical System.”*

Current situation

Currently, the concepts, classifications, variables and data collection instruments subsystems and a prototype for storage of methodological documents have already been implemented. All these subsystems are available on the organisation’s intranet. The concepts, classifications, methodological documents and data collection instruments are available on the Official Statistics website. The indicators and associated metadata disseminated on the website are registered in the variables system. Metadata on variables are available in Portuguese and English and the concepts and classifications are currently being translated.

In 2008-2012, the statistical metadata system will be adjusted so that metadata can be added to the data throughout the life cycle of surveys. The following actions are planned:

- A methodological documentation subsystem will be implemented to document any statistical object needing to be documented with any type of report, with the possibility of reusing metadata existing in the system in another context. In the initial phase, this system will cover methodological documents from surveys and reports in the SDMX format. In the second phase, a standardised quality report format will be implemented.
- The concepts subsystem will be reformulated so that the statistical concepts can be organised into conceptual systems and made available according to this criterion.
- Implementation of a structure that can be used to organise and relate object classes (statistical units and populations) in the variables subsystem.

2. STATISTICAL METADATA SYSTEMS AND THE STATISTICAL CYCLE

2.1 Statistical business process cycle The life cycle of primary statistical operations is the subject of the Statistical Production Procedures Handbook (in the approval stage), as shown in Figure 2.

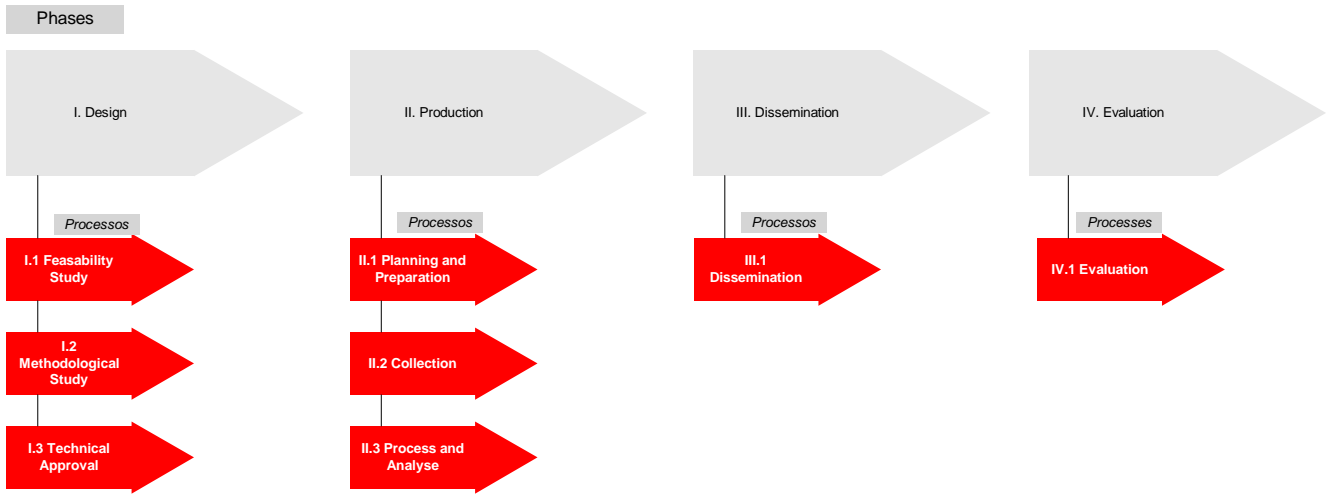


Fig.2. Phases and processes of the life cycle of primary survey

The processes are divided into sub-processes and tasks, as shown in the table below, establishing the relationship between the life cycle to be approved at METIS meetings and the life cycle used at the SP.

The table below (Figure 3) details only the processes and tasks needed to compare the two life cycles.

SSLC	Phase	Process	Sub-process/ Task	Responsibility	Documentation	File
	I. Design					
1. Need		I.1 Feasibility study		SPD	Feasibility study	
		I.2 Methodological study			Methodological document (preliminary)	
2. Develop and Design			I.2.1 Methodological study planning	SPD		
2. Develop and Design			I.2.2 Define survey contents	SPD		
2. Develop and Design			I.2.3 Define the sample	MISD		
2. Develop and Design			I.2.4 Define data collection methods	SPD		
			I.2.5 Design collection instruments and methodology	SPD		
2. Develop and Design			I.2.5.1 Design initial version of collection instrument and filling instructions			
3. Build			I.2.5.2 Test collection instrument			
3. Build			I.2.5.3 Build final version of collection instrument and filling instructions			
2. Develop and Design			I.2.6 Define quality control	SPD / MISD		
2. Develop and Design			I.2.7 Define process and analysis methods	SPD / MISD		
2. Develop and Design			I.2.8 Define dissemination products	SPD /DU		
2. Develop and Design			I.2.9 Define the requirements for application components	SPD		
3. Build			I.2.10 Manage field test/ pilot	SPD/ MISD / DCD		

SSLC	Phase	Process	Sub-Process/Task	Responsibility	Documentation	File
			I.2.11 Define survey budget	SPD		
		I.3 Technical approval				
			I.3. Technical certification	MISD	Methodological document (final)	
			I.3.2 Approval	Management		
	II. Production					
		II.1 Survey planning and preparation				
			II.1.1 Planning the survey	SPD / DCD	- Technical form for the data collection process - Data collection procedures handbook	
4. Collect			II.1.2 Methods development	MISD / SPD / DCD		
3. Build			II.1.3 Build support applications	MISD / SPD		
4. Collect			II.1.4 Training	SPD		
		II.2 Collection			Quality report	
4. Collect			II.2.1 Data collection preparation	DCD		
			II.2.2 Data collection	DCD		
4. Collect			II.2.2.1 Control collection tasks	DCD		
4. Collect			II.2.2.2 Control the reception of response	DCD		
4. Collect 5. Process			II.2.2.3 Analyse, classify, code and data capture	DCD		
5. Process			II.2.2.4 Validate data	DCD		

SSLC	Phase	Process	Sub-Process/Task	Responsibility	Documentation	File
5. Process			II.2.2.5 Test data coherence	DCD		
5. Process			II.2.2.6 Create microdata file from data collection	DCD	File documentation	Microdata collected
4. Collect			II.2.3 Manage non-response	DCD		
4. Collect			II.2.4 Manage providers	DCD		
4. Collect			II.2.5 Data collection quality control	DCD	Quality survey report	
5. Process			II.2.6 Update registers	DCD / MISD		
4. Collect			II.2.7 Update sample	MISD		
		II.3 Process and analyse				
5. Process			II .3.1 Statistical analyse on data coherence	SPD		
5. Process			II.3.3 Non-response imputation	SPD		
5. Process			II.3.4 Estimation and sampling errors	MISD / SPD		
6. Analyse			II.3.6 Control the data results quality	SPD	Quality report	
6. Analyse			II.3.7 Prepare microdata for dissemination	SPD	File documentation	Microdata for dissemination
6. Analyse			II.3.8 Analyse data	SPD		
7. Disseminate	III. Dissemination				- Methodological document (current version) - SDMX (Report)	
			Prepare indicators to insert in DDB	SPD	Insert the metadata associated to indicators (Variables)	Macrodata for dissemination
9. Evaluate	IV. Evaluate				Quality report	

Fig. 3. Correspondence between the processes and tasks of two life cycles (UNECE/EUROSTAT/OCDE and SP)

Where:

SPD – Statistical Production Departments:

Economic Statistics Department

Demographic and Social Statistics Department

National Accounts Department

DCD – Data Collection Department

MISD– Methodology and Information System Department

DU – Dissemination Unit

2.2 Current system(s)

The Integrated Metadata System

The Integrated Metadata System is constituted by several subsystems: Concepts, Statistical Classifications, Statistical Sources (including the components: Methodological Documents, Data Collection Instruments, and in future Administrative Sources and Questions) and Variables.

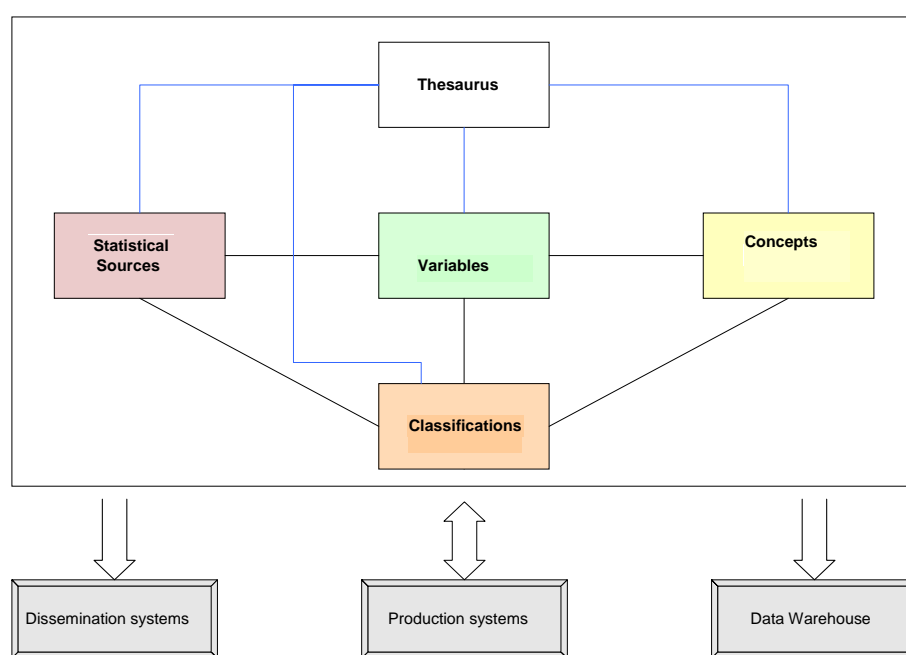


Fig. 4. Macro Architecture of the Integrated Metadata System

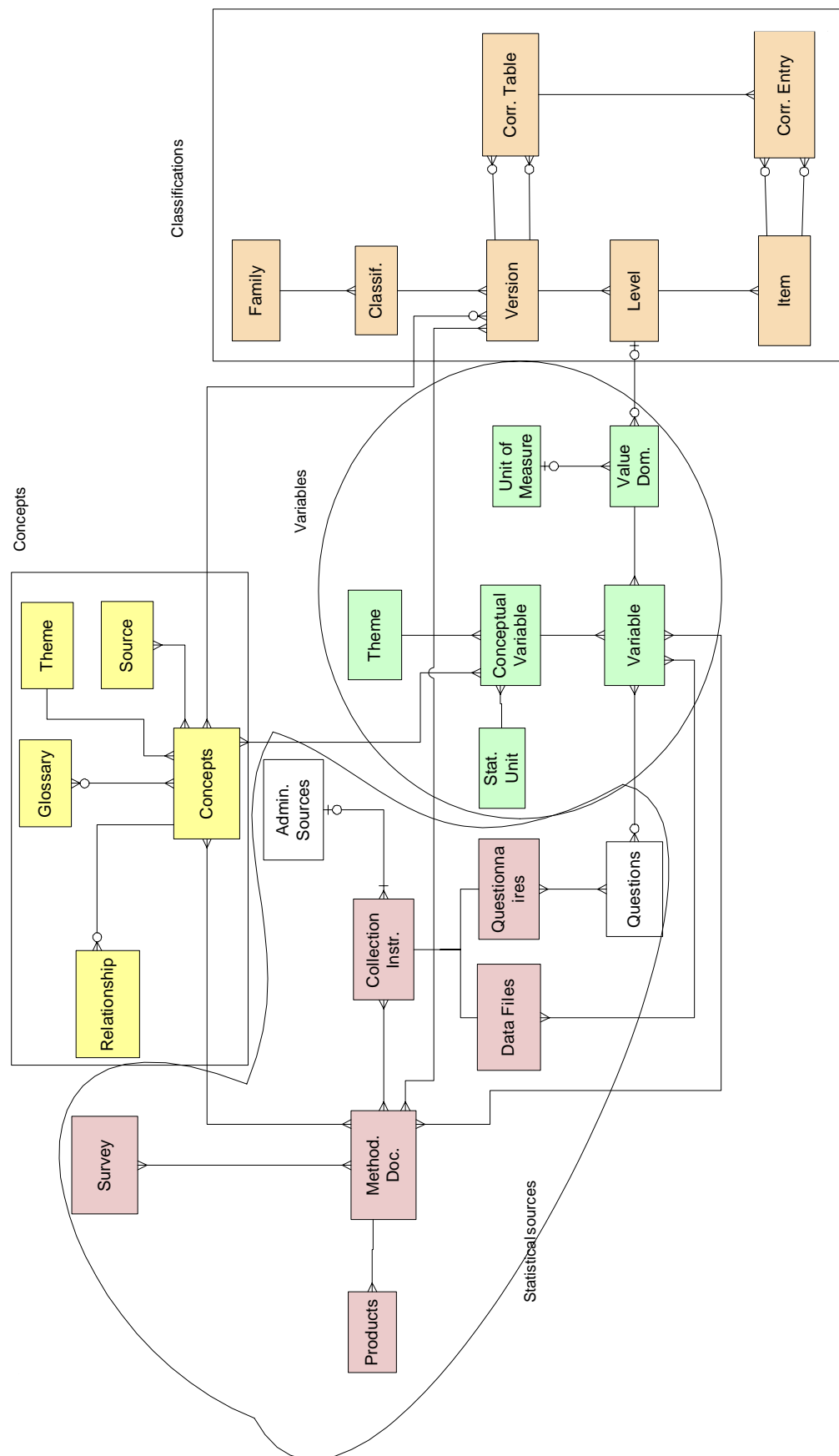


Fig. 5. Conceptual Model of the Statistical Metadata System

Purposes of the system

The main purposes of the integrated statistical metadata system are:

- To support the whole *life cycle* of surveys;
- To act as a *central repository* for statistical metadata serving as a source for other databases that support: design, production, dissemination of statistics and management;
- To establish *terminology* for statistical metadata;
- To constitute an *instrument for statistical harmonisation and coordination* of the NSS, standardising the documentation of surveys, among other elements;
- To implement a *homogeneous environment* for its technological infrastructure.

Concepts subsystem

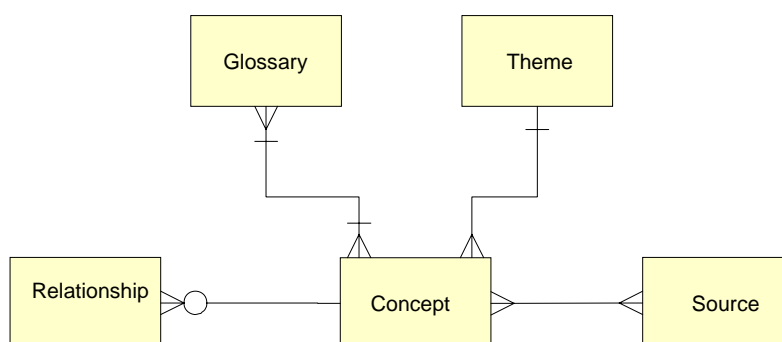


Fig. 6. Conceptual Model of the Concepts Subsystem

Main entity:

Concept – unit of knowledge created by a unique combination of characteristics (ISO 1087-1:2000, *Terminology work -- Vocabulary -- Part 1: Theory and application*).

The concepts and definitions recorded in the database are classified by subject area and organised in glossaries. Each glossary corresponds to a theme in the Official Statistics website.

The main attributes of the concepts are: code, name, definition, notes on the definition and source. Other attributes are required for the management of the system, such as status (proposed, in use, SC-approved), dates on which it was proposed, came into use and was approved by the SC. It is possible to establish a relationship between two concepts. Of these, synonymy and homonymy have already been implemented.

There is a generic glossary of concepts used throughout statistical activity entitled “Metadata Terminology” and a list of abbreviations and acronyms used in the documentation of surveys.

There is a plan to enlarge the system so that other types of relationship can be implemented which enable us to view the concepts of a particular area in the form of a conceptual system.

As a result of the integration of the different subsystems, the detail page of each concept shows its use in methodological documents, classifications and variables.

The concepts are available on the Official Statistics website, with access from the home page, and are searchable by alphabetical order in each glossary. An advanced search was implemented with the possibility of the combination of more than one search criterion.

It is in course the translation to English, of the concepts registered in the database. 20% of the concepts are, already, available in English.

Classifications subsystem

The conceptual model of the classifications subsystem was developed on the basis of the Neuchâtel model, a simplified version of which is shown in Figure 7.

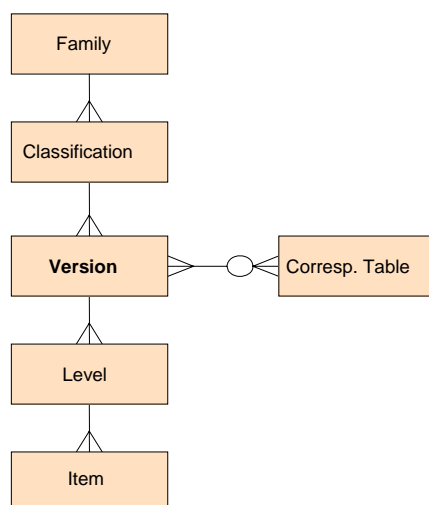


Fig. 7. Conceptual model of the Classifications Subsystem

The main purposes of this subsystem are:

- To constitute a reference for the NSS on national, EU and international nomenclatures and classifications used in statistics;
- To constitute an instrument to harmonisation and coordination for the statistical information;
- To constitute a management tool for nomenclatures and classifications.

Essentially, it provides access to three different types of information:

- National and international classifications and their description;
- Code lists (other grouping types);
- Correspondence tables.

Main entities:

Classification family – comprise a number of classifications, which are related from a certain point of view (e.g. products, economic activities, countries, etc.)

Classification - describes the ensemble of one or several consecutive classification versions. It is a "name" which serves as an umbrella for the classification version(s).

Classification version – a structured list of discrete, exhaustive, mutually exclusive categories defined by codes and designations intended to typify all units of a certain population in relation to a defined property. A classification version has a certain normative status and is valid for a given period of time.

Classification level – a level of aggregation of a classification; all categories at the same level have the same code structure. In a hierarchical classification the items of each level but the highest (most aggregated) level are aggregated to the nearest higher level. A linear classification has only one level.

Classification item - represents a category at a certain level within a classification version or variant.

Correspondence table – relationship between different versions of the same classification or between versions of different classification

This subsystem allows:

- To consult and export classification versions, respective correspondence tables and indexes, when they exist;
- To consult a set of normalised attributes that characterise each classification version;
- To consult other specific and relevant attributes in determined classification versions;
- To consult documentation related with each classification version;
- To consult variants of a classification version;
- To consult, by date, “floating” classification versions.

The classifications are accessible through the home page of the Official Statistics website.

Variables subsystem

The conceptual model is based on international standard ISO/ IEC 11179, “Information Technology – Specification and Standardization of Data Elements” (Figure 8).

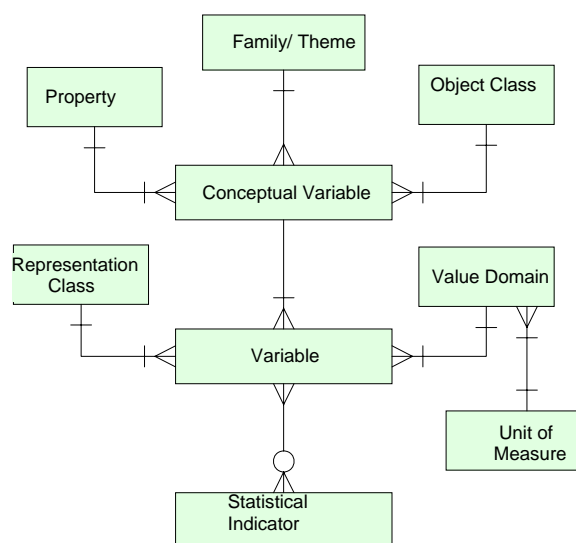


Fig. 8. Conceptual Subsystem of the Variables Subsystem

The variables subsystem provides a database of variables standardised and harmonised with their respective concepts, classifications, explanatory notes and calculation formulae.

The main purposes of the variables subsystem are:

- To support the questionnaire and survey design;
- To improve statistical coordination;
- To support the dissemination of statistical data;
- To assist the definition of normalized and/or harmonized variables;
- To promote comparability of data by using normalized variables.

Main entities:

Variables family – a classification for variables in general to facilitate the search for

variables in the system.

Property – characteristic or attribute common to all members of an object class; a property is a concept.

Objects class - a set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning whose properties and behaviour follow the same rule.

Object classes in this subsystem are:

- Statistical units;
- Populations.

Conceptual variable – a property of an object class described independently from any particular representation.

Representation class – a component of the definition of the variable indicating the type of data it represents (code, ratio, quantity, etc).

Value domain - a set of permissible values and their associated meanings. The value domains may be:

- Categorical (or discrete);
- Continuous;
- Text.

Variable – the smallest identifiable unit of data in this subsystem for which a value domain, a unit of measure, versions, permissible values can be specified.

Statistical indicator – a data element that represents statistical data for a specified time, place, and other characteristics. It consists of a cross-reference between an aggregate variable and classification variables called dimensions. Each indicator has at least two dimensions: time and geography.

Example: Resident population by place of residence, sex and age group.

At present, all the statistical indicators disseminated on the Official Statistics website, are registered in this subsystem, with complete metadata in Portuguese and English.

Data collection instruments subsystem

The data collection instruments subsystem stores and publishes in user interface, all the questionnaires (files still in preparation) that represent an instrument of reference on data used in NSS surveys. Images of questionnaires are available too, as well as some of its characteristics as: frequency and the variables that it observes.

The main purposes of the collection tool subsystem are:

- To constitute a repository of data collection instruments used in NSS surveys;
- To constitute a management tool for collection instruments.

There are basically two types of statistical data collection instruments:

- Questionnaires;
- Files.

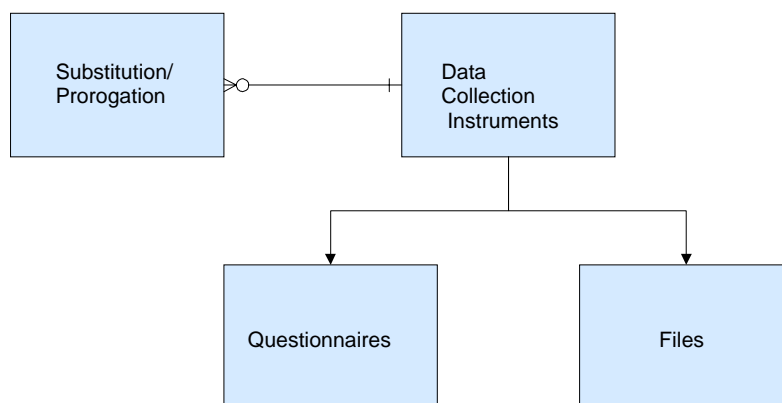


Fig. 9. Conceptual Model of the Data Collection Instruments Subsystem

This subsystem makes it possible to:

- Consult and manage questionnaires and files;
- Consult and manage the history of different collection instruments;
- View their images and layouts;
- Find out how they are used in methodological documents;
- Find out what variables they observe.

Main entities:

Data collection instrument – the means of transporting information from source to destination.

Questionnaire – an identifiable instrument containing questions designed to collect data from respondents.

File – a set of structured information in digital form.

The register of a data collection instrument is the final act of the technical certification process of a survey and guarantees the overall quality of the survey's object.

Data collection instruments are given a *registration number* and *period of validity*, whenever:

- It is a collection instrument in a new survey;
- There have been changes to the *content* of a collection instrument in a routine survey resulting from:
 - Inclusion or exclusion of variables;
 - Changes to questions;
 - Change on the name of survey.

Methodological document subsystem

This is the core subsystem in statistical production and the one that interacts most directly with the life cycle of surveys: in the *design phase*, surveys managers define the methodologies, concepts and classifications to be used, questionnaires and their connection to the list of observation variables and definition of data for dissemination.

The methodological documents of surveys have a normalized format in order to facilitate and increase their usability (Figure 10). This standard format was approved by the Statistical Council to document all the surveys in the NSS.

I. General characterization	Code/ Version /Approval date
	Code SIGINE
	Name
	Statistical activity / Statistical domain
	Purpose
	Description
	Responsible entity
	Relation with EUROSTAT/ other entities
	Financing
	Legal frame
	Type of survey
	Type of data source
	Obligatoriness
	Frequency
	Geographical scope
	Users
	Begin/ end date
	Products
II. Methodological characterization	Target population
	Frame
	Sample unit
	Observation unit
	Sample design
	Questionnaire design
	Data collection
	Imputation
	Estimation
	Time series
	Confidentiality
	Disclosure control
	Quality evaluation
	National and international recommendations
III. Concepts	
IV: Classifications	
V. Variables	Observation variable
	Derived variable
	Indicators
VI. Data collection instruments	Questionnaires
	Files
VII. Abbreviation and acronym	
VIII. Bibliography	

Fig. 10. Standard format of the methodological document

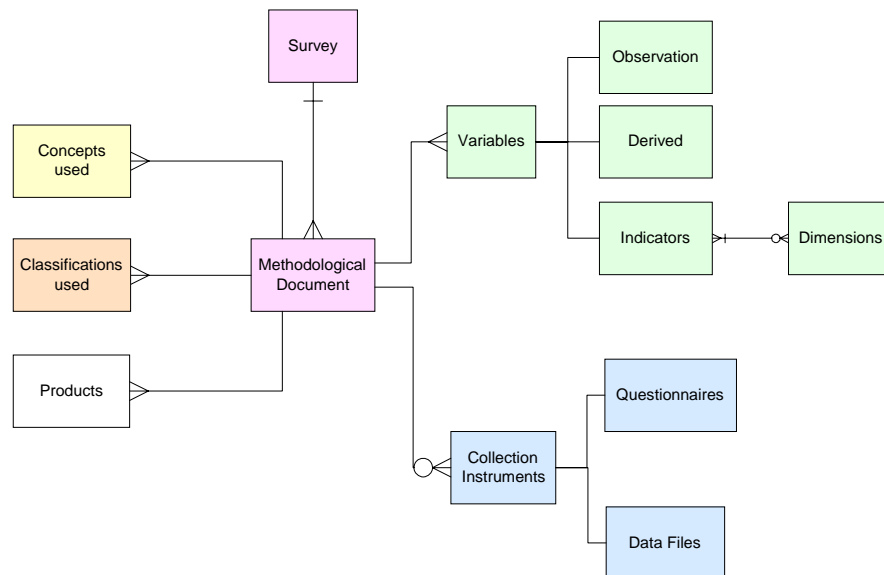


Fig. 11. Conceptual Model of the Methodological Document Subsystem

Main entities:

Survey – a statistical activity belonging to a predefined statistical method and involving the collection, processing, refinement, analysis, study and dissemination of data on the characteristics of a population. Four basic types of surveys are considered: sample survey, census, analytical study and statistical study.

Questionnaire – an identifiable instrument containing questions designed to collect data from respondents.

Method – a structured approach to solving a problem.

This entity contains the characterisation of methods of collecting data, designing samples, allocating answers and estimating and calculating errors, among others.

Universe – all the elements (people, entities, objects or events) with a given common characteristic.

Sampling frame – a list of units belonging to a given population used to select samples. Sampling frame must be characterised by the design methodology, updating system and quality control.

Sample – subset in a population or universe.

Phases/ Processes Metadata System	Planning	Design	Operation			Dissemination	Evaluation
			Preparation	Data Collection	Processing		
Concepts		I,C		C		I,C	
Classification		I,C		C		I,C	
Methodological Documentation	C	I	C	C		C	
Data Collection Instruments		C		C		C	
Variables		I,C	C	C	C	I,C	

Fig. 12. Interaction between metadata subsystems and the life cycle of statistical operations

Where: I – Inserted

C – Consulted

2.3 Costs and Benefits

All the system was implemented in-house, with one exception: the prototype system for the management of methodological documents that was implemented in outsourcing. This way we reduce costs and the maintenance of the system is easier. On the other hand, as IT technicians are not enough to all the agency needs, implementation time increases.

Since 2003, three IT technicians/year, on average, have been assigned to the development of the statistical metadata system.

2.4 Implementation strategy

The different subsystems, of which the general lines had been presented and approved by the Board and the Council of Directors in May 2002, were then detailed and implemented. Each one's information requirements, user interfaces, uploading and updating procedures, rules on content and plans for the use of existing information were defined in the details of these subsystems.

Implementation priorities are defined on the basis of the institution's needs.

After the general lines were approved for the metadata system mentioned in point 1 "Metadata Strategy", it was implemented as follows:

- We studied the implementation of metadata systems by other statistical institutes, such as that of Statistics Canada (2002-2004).
- We defined the system's conceptual model to integrate its different components.
- An existing subsystem of statistical concepts implemented in 1994 was initially thought to be appropriate.
- We implemented a classification subsystem (2003-2006).
- We defined a standard format for methodological documents in surveys (2003-2004), which was approved by the Statistical Council for documenting all NSS surveys (2005).
- We implemented a prototype subsystem to store methodological documents (2003-2004).
- We reformulated a questionnaire management subsystem implemented in 1997 (2006-2007).
- We implemented the variables subsystem (2004-2007).

3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL BUSINESS PROCESS

3.1 Metadata Classification

One way of regarding the role that metadata can play is to identify their function in the different statistical processes and respective tasks:

Statistical metadata functions:

- Contextualising data and supporting their dissemination and re-use;
- Giving information on the quality of the data provided;
- Harmonising concepts, classifications and questions, promoting comparability of information;
- Documenting production processes.

Statistical metadata include a wide range of attributes. We can therefore consider another level of classification:

- Survey metadata – in this category we consider all metadata for characterising the survey and schedules required for the planning and dissemination of data (attributes of Chapter I of the methodological document - general characterisation of the survey - and the schedules for data collection and dissemination of results).
- Methodological metadata – a description of the methods supporting the processes associated to the survey (attributes of Chapter II of the methodological document - methodological characterisation of the survey).
- Definitional metadata – includes the concepts, classifications, definitions of variables and questionnaires used.

- Quality metadata – includes all the attributes in the quality reports and indicators defining the quality of a survey.

System metadata - information required by operating systems and programs to function properly. It is destined to supply the information on the physical representation of data and other technological aspects and to support exchanges of information between systems.

3.2 Metadata used/ created at each phase

In the SP's metadata system, the most active phases in the insertion of metadata into the system are design and dissemination. In the operation phase, the data collection process is the one involving the most collection and use of metadata. This idea is based on an analysis of the electronic collection system (WebInq) and projects for the “universes and samples”, “surveys process management”, “statistical burden indicators” and “household survey questionnaire systems”.

WebInq

WebInq is an online service available on the Official Statistics website for electronic data collection. It allows respondents to answer SP surveys in different ways:

- Filling in an electronic form online;
- Filling in XLS (Excel) files and sending them by email;
- Uploading XML files.

For each survey whose data can be collected in this system, we have described some of the characteristics included in its methodological document and show an image of the questionnaire in the data collection instruments system.

The information from the methodological document visible on WebInq comprises: description, objectives, legal framework, type of survey, geographical scope, date reference period, data collection period, concepts and classifications used. The surveys are identified in the system by the survey code used in the metadata system.

Universes and Samples management system

This system is in its initial implementation phase and its purpose is the integrated management of an annual universe frame to support all the surveys based on the “enterprise” statistical unit. Two other sub-universe frames are created on the basis of this universe, one to support short term surveys and the other to support structural surveys. The sample frame and the samples are selected from these sub-universes. The entities making up this system are: universe frame, sub-universe frame, sample frame, sample and stratum. The attributes of the survey entity that are not featured in the methodological document, but are required by the entity, will also be defined in the system.

For a survey to be processed in this system, its methodological document must be registered in the metadata system.

In order to characterise the samples in a survey, certain attributes must be defined and identified: the universe with which it is associated, the names of the stratum variables, the names of the changeable variables, frequency, possibility of replacements and the replacement method and associated questionnaires, among others.

In this system, surveys are identified by the code used in the metadata system.

Collection process management system

This system is in the development phase and will provide transversal support to surveys, to the different components of the data collection processes for self-completion surveys:

- Control of the data collection operation;
- Despatch of survey;

- Receipt of responses;
- Preparation of auxiliary charts for controlling responses.

This system interacts with:

- “Universes and samples management system” - importing samples for the despatch of the survey and updating samples on the basis of information on replacements;
- Planning and control system - importing established schedules;
- Metadata system - its gets information on the characteristics and code of survey.

Statistical burden indicators

This system is in the planning stages and, when implemented, will be a tool for analysing statistical burden and the enterprise response rate. From the metadata system, we expect the use of the survey code and name, registration number and questionnaire name, association of questionnaires with surveys, frequency and variables observed in the questionnaires.

Dissemination Database

This database of indicators supports the dissemination of statistics. The data disseminated are accompanied by their metadata. The statistical metadata come from the metadata system and no indicator can be provided without being recorded in the variables subsystem of the metadata system.

The variables defining an indicator (variable measure and its dimensions) are recorded in the variables subsystem along with a cross-reference between the variables that defines the indicator. There are rules on naming variables and indicators. The value domains of the dimensions are also recorded in the classification subsystem and the concepts measured by the variables in the concept subsystem. Each indicator is given a code which links the two systems.

The metadata attributes provided for each indicator are its name, frequency, source, unit of measure, associated concepts, definition, formula and other contextual information.

The table below shows which metadata entities are inserted (I)/ updated (U)/ consulted (C) in the different phases, processes or documents produced in the life cycle of statistical operations.

Phase/Process/Document	Annual planning	Design			Operation								Dissemination		Evaluation
					Data Collection							Process and Analyse			
		FS	MD	Survey Certification	Populations and Samples	WebInq	Data Collection Technical Report	Household Surveys	DC Proc. Mngt.	Statistical Burden Evaluation	Microdata (1st. Inst.) File Doc.	Microdata (2nd InRt.) File Doc.	DDB	SDMX	
MD - Methodological document		I	I, U, C					C							
S - Survey	I	I, U, C	I, U, C		C	C	C	U, C	C	C				C	I, C
PD - Product	I	U	I, U												
CNC - Concept			I, C		C						C	C	I, C		
CL - Classification		I	I, U, C			C					C	C	I, C		
VAR - Variable			I, C	C				U		C	U	U, C	I, U, C	I, C	
DI - Disseminated information			I, C										I, C	I, C	I, C
DCI - Data collection instrument		C	C	I	C	C	C	C		C					
CO - Collection occurrence	I						I	I, U, C	I						I
SID – Survey instance data	I		C					C	U, C	C					I, C
US - Univers/ Sample					I, C		I			C				I	
AG - Agency										C				I	

Fig. 13. Metadata entities and the life cycle of statistical operations

Please note that the unimplemented systems and documents in this table are shadowed to distinguish them from the ones that have been implemented.

3.3 Metadata relevant to other business processes

There are close links between the activity planning, human resources planning and budget subsystems. In the second half of each year, departments prepare the activity plan for the following year. This plan lays out all the surveys to be undertaken in that year in national statistical production. When the unit heads draw up the plan, they define the schedules of surveys and allocate human resources on a person-hour basis. Personnel costs are calculated automatically on the basis of this allocation. The activity plan includes some characteristics of the surveys into the metadata system, such as their code, name, frequency, the responsible entity, type of survey, observation unit and sample size. Activities are given a code for analytical accounting. This is the code used to draw up the budget. The plan also refers to this code and the methodological document saves it so that the three systems can be linked.

4. SYSTEMS AND DESIGN ISSUES

4.1 IT Architecture

Each subsystem in the integrated metadata system has a similar architecture: a database, two Web applications (one for consultation and the other for management) and a view that provides metadata to be reused by other systems.

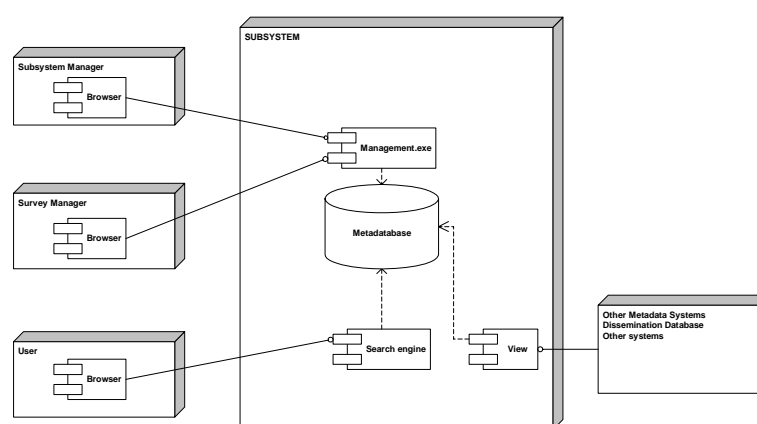


Fig. 14. IT Architecture

Management was designed to be decentralised with central coordination. The management application therefore implements two profiles: the *subsystem manager* and the *survey manager*. There is a generic profile for consultation.

4.2 Metadata Management Tools

The subsystem management applications in the integrated statistical metadata system were developed with the same computer infrastructures as those supporting all the SP's information systems.

Operating system

The tools at users' disposal are:

- The IIS servers and databases using Microsoft operating systems.

Network

The network architecture is based on open protocols and industrial standards and is comprised of local area networks (LAN) and wide area networks (WAN).

Applications

The applications supporting the metadata system are Web applications developed with the ".NET" platform. All the subsystems have a bilingual consultation application and a management application.

Storage

The huge amount of information produced or collected into the system requires an appropriately sized database. The associated databases (relational databases) are developed in Microsoft SQL Server so that it is easier to integrate with the production and

dissemination systems and the data warehouse.

Servers






	Server	Characteristics	Observations
INTERNET		Compaq ProLiant ML370 Microsoft Windows Server 2003, Standard Edition Service Pack 2 2 X Pentium III /800Mhz 2 GB RAM 72 GB Logical Disk Space	Web Sites, like: "metaweb.ine.pt", etc
Development Environments		Compaq ProLiant ML370 Microsoft Windows Server 2003, Standard Edition Service Pack 1 2 X Pentium III /733Mhz 1 GB RAM 64 GB Logical Disk Space	WEB Development
	 SQL	Compaq ProLiant ML570 Microsoft Windows 2000 Server, Service Pack 4 2 X Pentium III Xeon /700Mhz 1 GB RAM 72 GB Logical Disk Space	SQL Server 2000
SQL Production	 SQL	Compaq ProLiant DL760 Microsoft Windows 2000 Advanced Server Service Pack 4 4 X Pentium III Xeon /900Mhz 4 GB RAM 720 GB Logical Disk	SQL Server 2000
INTRANET		HP ProLiant DL380 G2 Microsoft Windows Server 2003, Standard Edition 2 X Pentium III /1.40GHz 1 GB RAM 72 GB Logical Disk Space	Intranet "intranet.ine.pt", "metaweb.ine.pt", etc.

Fig. 15. Servers

4.3 Standards and formats

The standard formats used are Excel, CSV and PDF.

4.4 Version control and revisions

The metadata are not static and change very fast due to modifications, for example, in concepts, classifications, revisions, new business rules, new methods, among others. As a result, there are "in use" and "not in use" versions of the entities in the integrated statistical metadata system which has to be created, verified and controlled. The versions are managed strictly by the managers of each subsystem, as already mentioned, following the rules governing each subsystem and obeying the rules of integrity. The subsystems are systematically revised in order to check the existing functionalities and ensure the ongoing implementation of improvements in usability and flexibility.

4.5 Outsourcing versus in-house development

The metadata system has been developed and implemented almost exclusively by in-house specialists. The reasons for this decision were:

- The existence of resources with good technical training;
- Good in-house knowledge of our statistics;
- The reduction in costs of undertaking the project;
- Assurance of continuous system maintenance.

Only the prototype system for consulting and managing the methodological document (as we mentioned before) was developed under an agreement with a university. The final version of the methodological documentation system that is expected to replace the prototype will begin to be developed later in 2008.

5. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES

5.1 Overview of roles and responsibilities

The metadata system user profiles that interact with the life cycle of surveys are as follows:

Metadata system manager – This job has thus far been done by the metadata system manager, whose duties are:

- Coordinating managers of each subsystem in the system;
- Ensuring that the different subsystems' conceptual models are properly integrated;
- Defining the general harmonisation rules applicable to all subsystems, in cooperation with the subsystem managers;
- Planning training courses, subsystem revisions, etc. in cooperation with the subsystem managers.

Metadata subsystem manager (central metadata unit)

- Each metadata, concepts, classifications, variables, methodological documents and data collection instruments subsystem has a manager in the Metadata Unit who guarantees the application of standardisation and harmonisation rules in each subsystem. These managers hold talks whenever necessary to articulate coherence and integrity between the different subsystems. They also have discussions with the *survey managers* and the Dissemination Database manager.
- The *concepts subsystem* manager manages the concepts database, guarantees the application of the terminological rules in the formation of concepts (in the allocation of names to concepts and the construction of definitions), and decides in which thematic area and glossary each concept should be classified. S/he standardises the source of concepts under standard NP 405 and provides support for the SC working groups and Production Departments when drawing up new concepts and in the periodical revision of the concepts in each thematic area, organise the concepts into a conceptual system, for each thematic area, provides the translation of concepts to English and manages the system's decoding tables. Also s/he interacts with the IT technicians in the implementation and maintenance of the subsystem and prepares the necessary documentation for sending concepts to the SC for approval.
- The *classifications subsystem* manager manages the classifications database. S/he ensures that there are no redundant classifications in the subsystem and that the names and versions of classifications and rules for classification and coding are harmonised. S/he arranges for classifications registered in the subsystem to be translated into English and, whenever possible, into French. S/he manages the system's own decoding tables and interacts with the IT technicians in the implementation and maintenance of the subsystem. S/he holds meetings with the managers of each classification to ensure the coherence and harmonisation of the subsystem and prepares the necessary documentation for sending classifications to the SC for approval. S/he interacts with the managers of the different classifications.
- The *variables subsystem* manager ensures that the rules governing this

subsystem are obeyed, so s/he checks proposed properties, object classes, representation classes and value domains to make sure that they are not duplicated. S/he also ensures that the names given to variables abide by the subsystem's rules. After conducting these checks, s/he approves or rejects the variables proposed by the SMs. S/he manages the system's own decoding system and interacts with the IT technicians in the implementation and maintenance of the subsystem.

- The *methodological document subsystem* manager ensures that the format in Word of the document received from the Production Departments complies with the standard format approved by the SC and that the contents of each topic are in agreement with the expected content. At the moment, the subsystem is a prototype and not available to the SMs, so the metadata unit that enters them into the database and publishes them. In the final version of the subsystem, the SMs will enter the contents of methodological documents into the subsystem, where they will have "for approval" status until approved. The manager of this subsystem will change their status to "in force". S/he meets with the SMs to ensure the coherence and harmonisation of the contents of these documents and prepares the necessary documentation for sending surveys to the Board for approval.
- The *data collection instruments subsystem* manager registers data collection instruments (questionnaires or files) in the database and allocates them periods of validity as requested by the SMs. S/he manages the subsystem's own decoding tables and interacts with the IT technicians in the implementation and maintenance of the subsystem.

Survey manager (SM) – This title is given to the statisticians (subject matter) in charge of each survey or to experts appointed by them. Their responsibilities in the system are as follows:

- At the end of the year and on an annual basis, the SM enters the survey plan for the following year into the planning system.
- S/he drafts the feasibility study.

The design phase of a survey begins with a feasibility study that must be approved by the Board. This study is drafted by the SM.

- The SM proposes the concepts, classifications and variables to be used in the survey. As there are SC-approved concepts and classifications for use in the NSS, they are the ones that should always be used unless they are not suited to the survey in question, in which case appropriate concepts and classifications must be proposed. When the variables to be observed or disseminated in each survey are being defined, those already defined should be taken into account and reused whenever possible.
- S/he drafts the methodological document in accordance with the SC-approved format. The methodological characterisation of surveys is usually carried out by methodologists in close cooperation with the survey managers.
- At the end of the design phase, as set forth in the rules, the SM distributes the methodological document and its questionnaire(s) to the units that will use the information produced and to the planning, methodology, metadata and information systems units for their opinion. On the basis of these opinions, s/he makes any appropriate changes and responds to each opinion, indicating the suggestions that have and have not been included with explanations. S/he alters the methodological document and questionnaire appropriately and sends their final versions to the metadata unit for the questionnaire to be registered and the methodological document to be published as the version in force.

- S/he re-plans surveys whenever necessary.

Classification manager (central metadata group or subject matter)

Each *classification* has a unit and a specialist in charge of its content. Classifications that are used throughout all statistical activity are managed in the Metadata Unit, while classifications specific to one system are managed in their respective Production Department. It is this specialist who manages the content of the different versions of a classification for which s/he is responsible. S/he interacts with the classifications subsystem manager.

IT manager

This is the IT technician who coordinates the development of applications in all the metadata subsystems, ensures that they are included in the technical plan and maintains the subsystems.

Consultation

Access to consultation applications is open to any user.

Responsible of implementing the SDMX standard

This is a specialist who plays an active part in the Eurostat Task Force to revise the Content Oriented Guidelines. S/he is responsible for studying the SDMX standard so that the metadata system can be adapted to its requirements.

5.2 Metadata management team

The Methodology and Information System Department at the Statistics Portugal (SP) has a Metadata Unit. Its main duties are:

- Design, coordination of development and permanent management of all aspects of the NSS metadata system;
- Coordination of the technical approval process of Surveys;
- Management of classifications that are used throughout the NSS.

The *metadata subsystem managers* and some *classifications managers* belong to this unit. There are other *classifications managers* in the Production Departments.

The specialised team attached to this section comprises 17 technicians and the head unit:

- *Technicians* with a more general profile, who participate in devising and testing the different metadata subsystems, manage them and assist in-house and external entities in using the system; currently they are 12 technicians with this profile.
- *Nomenclaturists*, who normally have degrees in economics and study and devise national classifications, monitor EU and international work on studying and devising statistical classifications, assist in-house and external entities in using classifications and give expert opinions; currently they are 4 technicians with this profile.
- *Terminologists*, who have language and literature qualifications and belong to the concepts subsystem. They assist the Production Departments and SC working groups in designing conceptual systems, drafting definitions, arranging for the translation of concepts into English and giving expert opinions; currently the Metadata Unit has 1 technician with this profile.

Some of these specialists represent the SP in intra- and extra-community bodies and participate in statistical cooperation programmes.

The Metadata Unit does not have its own IT specialists. Experts from the Application Development Unit and also the Methodology and Information System Department provide IT services.

5.3 Training and knowledge management

The metadata system has been introduced to the SP in presentations of some of its systems, such as the classifications, methodological documents and variables systems. Some training courses on the classification system and data collection instruments (version 1) have been given to dissemination practitioners in order to answer users' questionnaires.

A presentation was also given of a project undertaken in 2006 with the Linguistics Centre at University Nova, Lisbon, defining a method for constructing conceptual systems. This new way of analysing concepts will mean that the concepts system will have to be altered to allow the necessary types of relationship to be defined so that concepts can be presented in this way.

In 2007, we began implementing a training plan in the metadata system. Two training courses were given on the variables system and the rest of the training plan is scheduled to be implemented in 2008, with the exception of the course on methodological documents, which will only take place in 2009. The plan includes not only training on consultation and management applications in the different subsystems but also on their underlying concepts, conceptual models and terminology. User manuals are being prepared for the training courses.

The intranet has a glossary of metadata terminology containing the concepts used by the integrated metadata system.

The current training plan that will be held every year consists of five courses:

1. Integrated metadata system

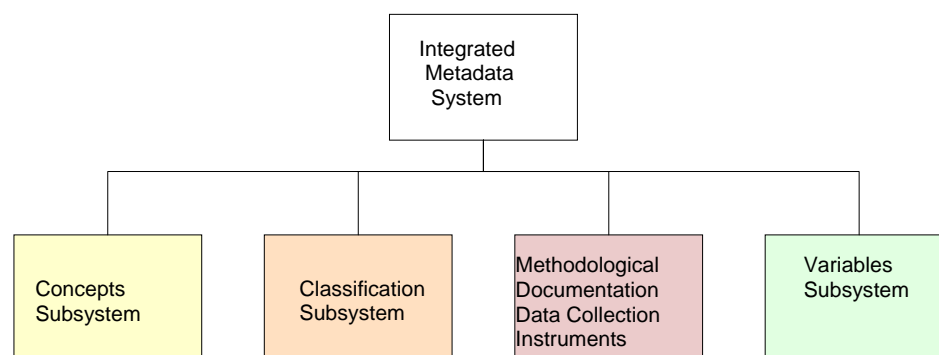


Fig. 16. Integrated Metadata System

This course is for all senior and assistant statisticians and includes the following subjects:

- The integrated metadata system's different components and its role as a technical statistical coordination tool;
- The basic concepts and terminology underlying this system;
- The role of each subsystem in the life cycle of surveys;
- Improving the statisticians' competences in the use of this system as a tool in designing the information subsystems in which they work;

2. Terminology

This course is designed to help senior statisticians develop the following competences:

- Designing conceptual systems;
- Preparing definitions;
- Identifying problematic definitions in the concepts database, that need a revision and improving according to the terminology criteria.

To achieve this, the course includes:

- Terminology theory;
- The notion of conceptual system and the application of a methodology in its design;
- Semantic relationships in terminology;
- Terminography;
- Definition.

3. Classifications systems

This course is for senior and assistant statisticians and includes the following subjects:

- The main international, EU and national classifications systems used in statistics, the classifications comprising them and the relationships between them;
- The approval process for classifications at different levels and harmonisation in the use of classifications in the NSS;
- Improving competences in the use of the classification subsystem.

4. Variables subsystem

This course is for senior and assistant statisticians and includes the following subjects:

- Recognition and designation of entities making up the conceptual model;
- The rules and principles in the name, standardisation and harmonisation of variables;
- How this subsystem relates to other subsystems in the integrated metadata system and other outside subsystems, such as the Dissemination Database and WebInq (electronic data collection);
- Use of the consultation applications and the survey manager (SM) profile of the management application;
- Defining indicators to be disseminated on the Official Statistics website.

5. Methodological documentation

This course is scheduled to begin in 2009 and is aimed at senior and assistant statisticians. Its subjects are:

- The SC-approved methodological document format for documenting the NSS's surveys;
- Use of the consultation applications and the survey manager (SM) profile of the management application.

The statisticians in the Metadata Unit participate in the above courses, EU-level courses and international conferences and practise high-level self-study.

5.4 Partnerships and cooperation

Statistics Canada's IMDB project was the main source of reference in developing the SP's integrated metadata system. When SP first began developing the system in 2003, some of its members visited Statistics Canada, where they followed a three-day programme set up by that agency:

Day 1:

- Brief, generic approach to the IMDB (Integrated Metadatabase) project;
- IMDB – Phase 2 (Description of a surveys, methods and quality metadata).

Day 2:

- COR (Common Object Repository);
- IMDB – Phase 3 (Defining variables).

Day 3:

- IMDB – Phase 3 (continued);
- Meeting with IT technicians involved in the project.

These were three days of highly useful work in which an in-depth analysis was made of some extremely relevant aspects that were addressed more generally in the project documentation available.

The following were also very important references for the definition of the system:

- The Corporate Metadata Repository model by Dan Gillman;
- The Neuchâtel model, which supports the classification subsystem;
- Documents on metadata systems, documentation and quality by Bo Sundgren:
 - *Documentation and Quality in Official Statistics*, Statistics Sweden, 2001;
 - *Objects and their Classifications, Relations, and Life Histories – as Reflected by Official Statistics*, Stockholm, Sweden: Statistics Sweden, 2004;
 - *Statistical Metadata – A tutorial*;
 - *The αβγτ-model: A theory of multidimensional structures of statistics*, Statistics Sweden, 2001;
 - *The Swedish Statistical Metadata System*, Statistics Sweden, 2000;
 - *The Contents of a Statistical System as a Whole*, Stockholm, Sweden: Statistics Sweden, 2004;

More recently, the Statistical Office of the Republic of Slovenia contacted the SP with a view to learning about its variables subsystem in detail. After analysing possible forms of cooperation, the SP provided the Slovenian agency with the subsystem's data model and has also responded to later requests.

As part of a statistical cooperation project with the Portuguese-speaking African countries, one of the Metadata Unit statisticians has been working with them on a project entitled “Classifications, Concepts and Nomenclatures”, which coordinates the five countries' economic classifications. Consultancy services have also been provided to a project to develop a common integrated economic nomenclature system for the five countries.

The SP also belongs to the Eurostat Metadata Task Force, which is analysing the main components of the SDMX Content-Oriented Guidelines: framework, inter-domain concepts and associated codification, vocabulary and statistical domains.

6. LESSONS LEARNED

6.1

We have certainly learned some lessons from the implementation of the integrated metadata system, which has been more systematic in the last six years, some because we have seen that our options have had a positive effect and others because we have realised the form they should have taken in order to be more successful. We are even making some changes in the formal circuits of some subsystems with a view to greater efficiency

and quality in the results obtained.

Involvement of the institution's top management was fundamental and the tie-in of the creation of documentation with formal and standardised procedures has been an excellent way of keeping documentation up to date at the SP.

- 6.2** Designing a metadata system not only requires considerable knowledge of statistical production, but also means leaving behind some habits acquired in this area. A great capacity for abstraction and tidy, integrated thinking is also necessary. An institution has specialists with all these capabilities but not always with all of them at the same time. The teams chosen to implement these systems must consist of specialists with different profiles among those mentioned, because they complement each other. The IT technicians who develop applications must participate from the start.
- 6.3** We believe that it is essential to develop prototype systems before final implementation. Prototyping is the best way to test a system's design, detect strong and weak points and come up with experience-based alternatives for the weak points. When designing a system like this, it is very hard to give an appropriate description of all its functions without prior experience. Even the workflow of procedures may need some adjustments.
- 6.4** Must be given training to statisticians, not only in the use of applications but also, and above all, about the concepts underlying the system and workflow of procedures. The introduction of the position of survey manager has fostered cooperation and dialogue between production, metadata and dissemination. The distribution of terminology associated with each metadata subsystem is having a beneficial effect at the SP as it encourages the use of a language common to all profiles using the system.
- 6.5** After the classification subsystem was made available to the general public, we began to receive some complaints about its usability and decided to conduct some usability tests. The test results showed us the difficulties that people experienced when using the system and we decided to redo some of the navigation in the consultation application. When we implement the methodological documentation subsystem, scheduled on the beginning of 2008, we have decided to conduct usability tests in the prototype phase of the consultation and publication application so that we do not need to redo any parts of the system after it goes into production.

*** END ***