

Statistics Portugal – Statistical Metadata System

Isabel Morgado

Summary

A major coordination tool to Statistics Portugal is the Integrated Statistical Metadata System. It supports both statistical production and dissemination of statistical data, and is built upon four main subsystems: Concepts, Classifications, Variables and Statistical Sources, with tightly coupled interrelationships.

The Metadata Unit, responsible for the central coordination of this system, has conceived and implemented it in accordance with survey managers, and deals with harmonization issues.

Statistical Council, which assures the coordination of the National Statistical System, approve concepts, classifications and other technical coordination tools. This paper aims to describe the Statistical Metadata System.

1. National Statistical System

The National Statistical System (NSS) consists of:

- Statistical Council (SC);
- Statistics Portugal - National Statistical Institute (SP);
- Central Bank – Bank of Portugal (BP);
- Regional Statistical Offices (Madeira, Azores);
- Entities producers of statistics by delegation of Statistics Portugal;

The Statistical Council (SC) is the state body that guides and coordinates the National Statistical System. Its mission includes, according to the Law 22/2008 of 13th May (Statistical Act):

- *“Define, every year, official surveys at a national level and those of regional interest, accordingly to the proposals of the statistical authorities;*
- *Approve technical instruments of statistical coordination, of mandatory use in the production of official statistics, promote their dissemination and use and propose to the Government their use in Public Administration;*
- *Approve and regulate standard procedures for the registration of data collection instruments submitted by statistical authorities and other sources that can be used for statistical purposes;*
- *Formulate recommendations on the definition of methodologies, concepts and statistical nomenclatures, to be used on administrative acts, to the production of official statistics and ensure their application;”*

All the above is carried out by the “Statistical Coordination” Standing Section (SCSS).

The job of Statistics Portugal, beyond the production of official statistics, is to supervise and make technical and scientific coordination of the NSS, taking into account the general guidelines produced by the Statistical Council. It may also delegate the production of official statistics 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.

1.1 Approval of concepts, classifications and methodological documentation

In these processes exists a strong interaction between SP and the Statistical Council. SP gathers all the information and prepares the documentation that is submitted to the SC for approval. The SP centralises the statistical concepts used in its own production and the delegated bodies’ statistical surveys in a database. These concepts are classified by subject areas and are loaded into the 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, with the collaboration of experts on a specific area or theme, to analyse them and recommend their approval to the SCSS. After the approval, their

status in the database is changed to “SC-approved concept”, and is of mandatory 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 SCSS 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.

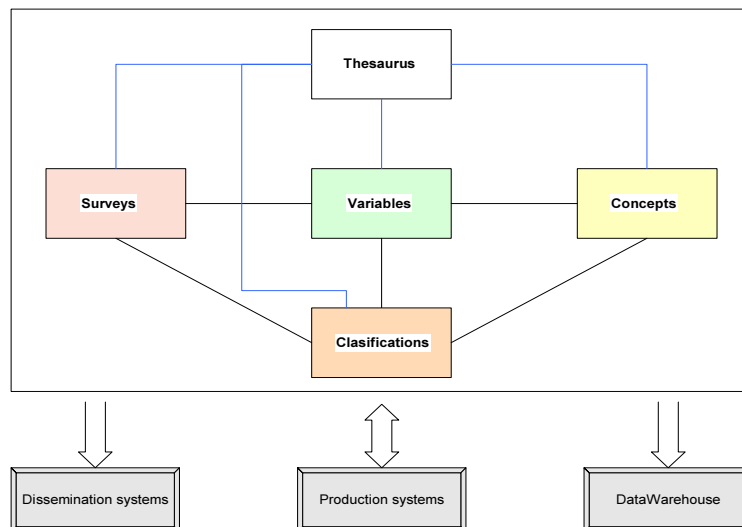
2. Statistical Metadata System

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

The main purposes of this system are:

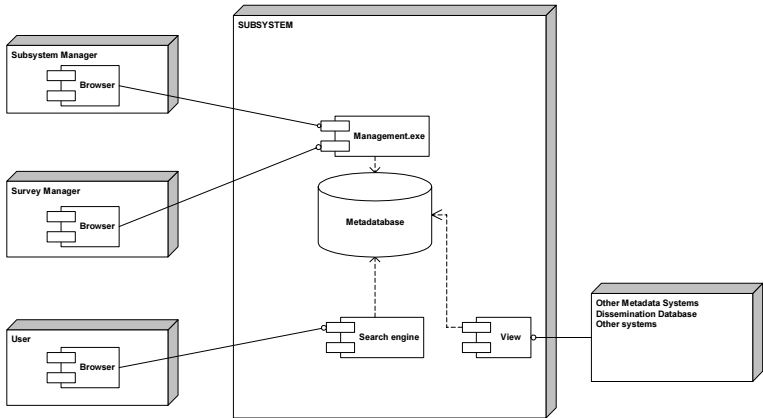
- To support the whole *life cycle* of surveys;
- To act as a *central repository* for statistical metadata serving as a source for other databases that support: design, production, dissemination of statistics and management;
- To establish *terminology* for statistical metadata;
- To be an *instrument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

Image 1 Macro architecture of the Integrated Metadata System



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.

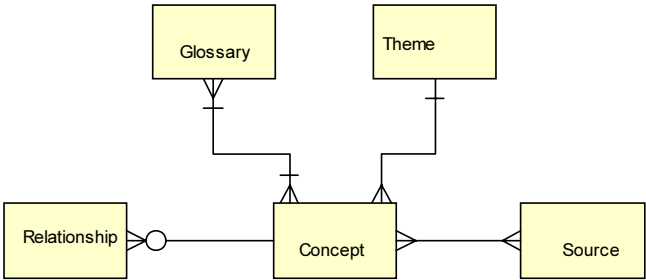
Image 2 IT Architecture



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.

2.1 Concepts subsystem

Image 3 Conceptual model of the Concepts Subsystem



A concept is a 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 Portal. 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, from which 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 scope of the system so that other types of relationships can be implemented enabling to view the concepts of a particular area as a conceptual system. Due to the integration of the different subsystems, the detail page of each concept shows its use in methodological documents, classifications and variables.

Image 4 Details of a concept

Detail	
	<input checked="" type="checkbox"/> In use 🇪🇸 🇬🇧
Code:	1477
Designation:	INACTIVE POPULATION
Validity:	Start: 29-04-2006
Thematic area:	LABOUR AND WAGES
Glossary(ies):	LABOUR MARKET
Definition:	All persons, regardless of the age, who, during the reference period, could not be considered to be economically active, i.e. who were neither 'employed' nor 'unemployed', nor on compulsory military service.
Notes:	-
Source(s):	Instituto Nacional de Estadística (INE)
Formula:	-
Synonym(s):	-
Homonym(s):	-
Historic:	INACTIVE PPOPULATION (24-05-1994) > INACTIVE POPULATION (29-04-2006)
Methodological Documents:	(138) - Labour force survey (Version 1.3) (136) - Labour force survey (Version 1.1)
Classifications:	-
Variables:	- Inactive population - Available inactive population

The concepts are available on the Official Statistics Portal, with access from the home page, and are searchable by alphabetical order in each glossary.

Image 5 List of the concepts of a glossary

The screenshot shows the 'Concepts - Statistical concepts - LABOUR MARKET' page. The search bar is set to 'LABOUR MARKET'. Below the search bar, there is an alphabetical index from A to Z. Under the letter 'A', the following concepts are listed with checkboxes:

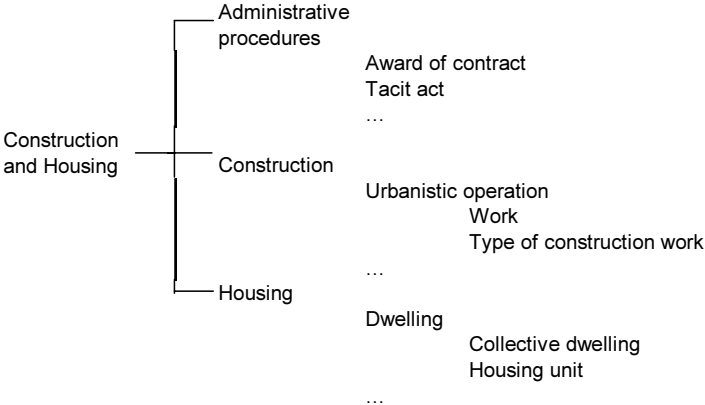
- ABSENTEEISM
- ACTIVITY RATE (population aged 15 years old and over)
- AN INDIVIDUAL'S PRIMARY OCCUPATION
- APPRENTICES AND TRAINEES
- AVAILABLE INACTIVE

Under the letter 'B', the following concept is listed with a checkbox:

- BASIC REMUNERATION

An advanced search was implemented with the possibility of merging more than one search criteria. It is ongoing the translation to English, of the concepts registered in the database, 50% of the concepts are, at the present, available in English. The concepts subsystem will be reformulated so that the statistical concepts can be organised into conceptual systems (Figure 10) and made available according to this criteria.

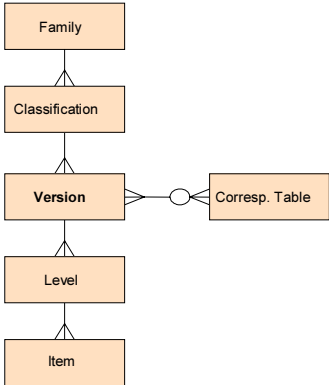
Image 6 Graphical representation of a conceptual system



2.2 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 Image 7.

Image 7 Conceptual model of the Classifications Subsystem

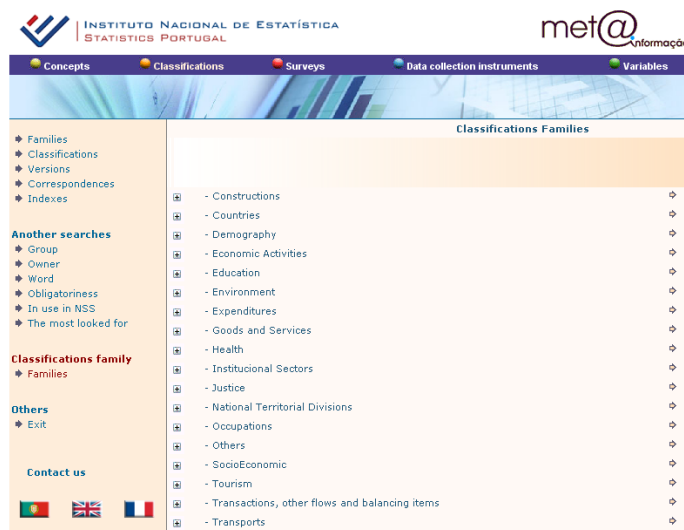


Essentially, it provides access to three different types of information:

- National and international classifications and their description;
- Code lists that are value domains of variables;
- Correspondence tables.

Related classifications are grouped in families, and a classification can have more than one version.

Image 8 List of Families of classifications



Accordingly to the Neuchâtel terminology a classification version is a “structured list of discrete, exhaustive and 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.

Image 9 The hierarchical structure of a 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 a specific version of each classification;
- To consult other specific and relevant attributes for a particular version;
- To consult documentation related with each version;
- To consult variants of a version;
- To consult “floating” versions, by date.

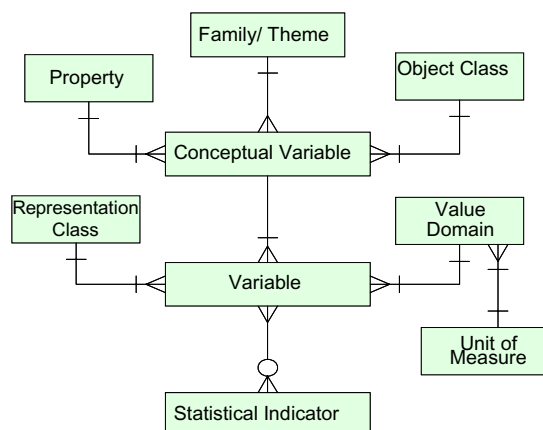
Fig. 8. Characterization of a classification version

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

2.3 Variables subsystem

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

Image 11 Conceptual model of the Variables Subsystem



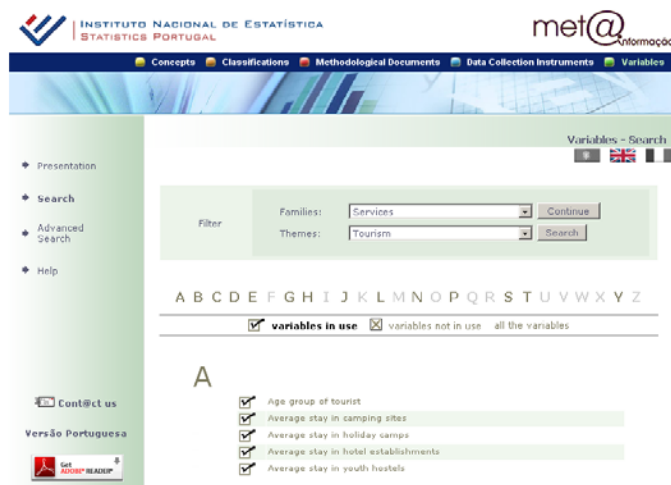
The variables subsystem provides a database of variables standardised and harmonised with their respective concepts, classifications, explanatory notes and calculation formulae.

The main purposes of the variables subsystem are:

- To support the questionnaire and survey design;
- To support the dissemination of statistical data;
- To assist the variables harmonization work;
- To improve statistical coordination;

Variables are classified in themes. Choosing a theme, a list of alphabetic ordered conceptual variables is provided.

Image 11 List of conceptual variables of a theme



Clicking on a conceptual variable, turns details accessible and the list of the variables that depend on it, clicking again on the name of the variable we access to it's details.

Image 13 Detail of a variable



A statistical indicator is a data element that represents statistical data for a specified time, place, and other characteristics. It is composed by several variables with different roles: a variable measure and several dimensions. Time and geography are mandatory dimensions. At present, all the statistical indicators disseminated on the Official Statistics Portal, are registered in this subsystem, with complete metadata in Portuguese and English.

Image 14 – Definition of a statistical indicator

Observação (0) Derivada (0) Medida (18) Dimensões

Selecione a Medida

3156|2007-04-26 > Alunos matriculados no ensino básico (N.º)
 3180|2007-04-27 > Alunos matriculados no ensino secundário (12º ano via de ensino - N.º)
 3181|2007-04-27 > Alunos matriculados no ensino secundário (Nível 3 dos cursos profissionais - N.º)
 3194|2007-04-27 > Alunos matriculados no ensino secundário (Artes visuais - N.º)
 3194|2007-04-27 > Alunos matriculados no ensino secundário (Cursos tecnológicos - N.º)

Cruzamentos: 2260

190|2005-08-03 > Período de referência dos dados
 3182|2007-04-27 > Localização geográfica (País - 2002) do estabelecimento de educação e ensino
 310|2005-01-01 > Sexo (Difusão - HM) do indivíduo
 3173|2007-04-27 > Nível (ensino básico) do estabelecimento de educação e ensino

Gerar designação do indicador

Indicador SVAR Cod: 2260

Designação Formal: Alunos matriculados no ensino básico (N.º)_Período de referência dos dados_Localização geográfica_Sexo_Nível

Designação Divulgação: Alunos matriculados no ensino básico (N.º) por Sexo e Nível (ensino básico)

PUPILS enrolled on basic education (No.) by Sex and Degree (basic education)

 Outro:

Registrar

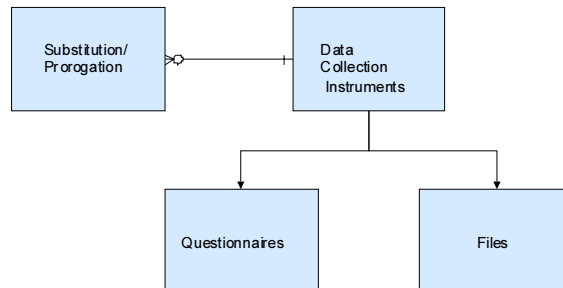
2.4 Data collection instruments subsystem

Data collection instruments are the means of transporting information from source to destination. The data collection instruments subsystem stores and publishes through a 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 such as frequency and the observation variables.

There are two main types of statistical data collection instruments:

- Questionnaires;
- Files.

Image 15 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.

Image 16 Filter to search Data Collection instruments

INSTITUTO NACIONAL DE ESTATÍSTICA
 STATISTICS PORTUGAL

met@ Informação

Concepts Classifications Methodological Documents **Data Collection Instruments** Variables

Data Collection Instruments - Search

It's necessary to restrict a characteristic/field in order to select the correspondent data collection instruments.

Research field: Geographic Coverage Continue

Value: Continente

In force Not in force All Instruments

Search Results:

Registration Number	Name	In Force Status
8830	Short-term statistic - Consumers monthly survey	<input checked="" type="checkbox"/>


Back to the previous display Export Back to the search formulary

Versão Portuguesa

Registration of a data collection instrument is the final step 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 routinized survey resulting from:
 - Inclusion or exclusion of variables;
 - Changes on questions;
 - Change on the name of survey.

Image 17 Detail of a Data Collection instrument

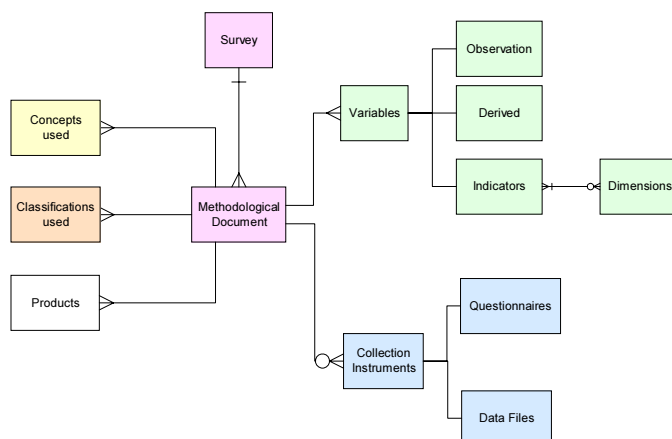
Registration Number:	8830			
Name (Instrument):	Short-term statistics - Consumers monthly survey			
Validity:	Start: 13-01-1998	End: 31-12-2006		
Geographic Coverage:	Continente			
Periodicity:	Monthly			
Type of Instrument:	Questionnaire			
Questionnaire:	 8830.pdf Type: application/pdf - Size: 87191 (bytes)			
Data Collection Method:	CAPI - Computer assisted personnel interview			
Respondent Entity:	-			
Statistical Observation Unit:	-			
Statistical Area:	-			
Responsible Entity:	Serviço de Análise de Conjuntura e Previsão			
Methodological Documents:	(62) Short-term statistics on business - Consumers survey (versão 1.0)			
Historic Information:	Procedures occurred	Date	Office note n°	Validity date
	Registration	13-01-1998		31-12-1998
	Prorogation	14-12-1998	1274/1998	31-12-1999
	Prorogation	23-11-1999	1141/1999	31-12-2000
	Prorogation	21-12-2000	953/2000	31-12-2001
	Prorogation	15-02-2002	037/2002	31-12-2002
	Prorogation	06-11-2002	193/2002	31-12-2003
	Prorogation	16-10-2003	112/2003	31-12-2004
	Prorogation	05-01-2005	004/2005	31-12-2005
	Prorogation	22-12-2005	349/2005	31-12-2006
Variables:	Age group of person Attended level of education of person Evaluation of change in prices over the last 12 months of person Evaluation of tax change in prices over the last 12 months (%) of person Evaluation of tax change in prices over the last 12 months of person Evaluation of the economic situation in the country over the last 12 months of person			

2.5 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*, survey 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 standard format in order to facilitate and increase their usability.

Image 18 Conceptual model of the Methodological Document Subsystem



This standard format was approved by the Statistical Council to support the documentation of all the surveys in the NSS and it is composed by 8 chapters:

- I – General characterization*
- II – Methodological characterization*
- III – Concepts*
- IV – Classifications*

- V – Variables*
- VI – Data collection instruments*
- VII – Abbreviations and acronyms*
- VIII – Bibliography*

In this context a *survey* is 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.

Image 19 Filter to search methodological documents

Methodological Documents - Search

Select the "fields" below in accordance with the options available.
Some contents only in Portuguese.

Themes: Labour market
Code: Version:

Word(s) to find:

All words Part of a word Any of these words

Search Restrictions: Themes = Labour market

Code	Name	Theme
(139)	Labour cost index (Version 1.2)	Labour market
(256)	Social Protection, Trade Unions and Employers Associations Statistics (Version 1.0)	Labour market

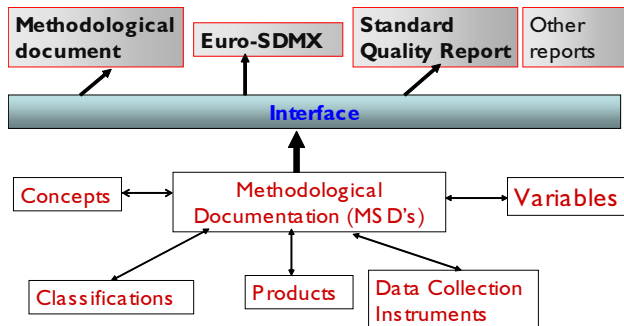
2 Records

The search is done applying a filter using theme, code, version or a string in the name of the survey. Clicking on the name, a PDF version of the document is shown. This documentation is available in Portuguese only.

A methodological documentation subsystem (Figure 9) 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.

Image 20 – Methodological documentation subsystem



3. Harmonization

The terminology used in the Organization depends very much on the statistical metadata system if it is prepared to support all the phases of a statistical activity from its design until dissemination.

Choose concepts already defined and stored in the system, code lists with a broad use stored as standard and variables already defined and available is a work that involves people responsible for the statistical metadata system management, subject matter statisticians and IT technicians.

Concerning the Concepts and Definitions subsystem, we are establishing some rules to be followed by all teams that deal with the analysis of concepts: rules regarding the term that designates the concept, rules to construct a definition, the difference between “definition” and “notes to the definition”. If all the teams work with the same rules, we believe that we’ll have a more harmonized concepts and definitions database.

The Classifications subsystem stores standard classifications and code lists that oftenly are variants of the standard classifications; establishing rules to name the code lists, giving the same code to the same category in different code lists, establishing the relationship between the code lists and the standard classifications they derive from, are some of the actions we do.

Another kind of harmonization we deal with is to make the names of variables and concepts that support them and the terminology used on the question that measure the variables, similar.

The variables subsystem has implemented a “naming convention” to the variables and to the statistical indicators for dissemination.

To reach a good result in applying all the harmonization rules described, a deep interaction with the subject matter statisticians is needed; it is very important that all the people involved on the task of defining statistical metadata be aware with the benefits of harmonization.