The Metadata System is composed of several integrated subsystems: Concepts, Statistical Classifications, Methodological Documents, Data Collection Instruments and Variables.

To the exception of the Data Collection instruments subsystem all other subsystems are available for direct consultation on the homepage of the Official Statistics website – Portal (http://www.ine.pt/).
The main purposes of the Metadata System are:

- To act as a central repository for statistical metadata to be used as source for other databases such as: design, production, data dissemination and management
- To settle terminology for statistical metadata
- To improve statistical literacy once it helps to better understand and interpret statistical data
- To implement an homogeneous environment to support its technological infrastructure
- To stand for an instrument for statistical harmonisation and coordination of the National Statistical System, standardising the documentation of surveys, among other elements

Methodological document

- To support the whole life cycle of surveys => thorough linked to the

Technical approval of surveys

The process of technical approval of surveys comprehends the following stages:

- For surveys on INE-PT responsibility:

  1. **1st Stage:** The preliminary methodological document and the questionnaire(s) produced in the methodological study (Design phase in the Business Process Model) by the production department are sent to the Units that intervene or use the survey results: such as, the Planning Unit, the Data Collection Department, the National Accounts Department and the Methodology and Information Systems Department (the coordinator of the process) for their opinion.

     **Deadline (for contributions):** 10 weekdays;

  2. **2nd Stage:** The Metadata Unit is part of the Methodology and Information Systems Department and in this Unit the methodological documentation manager as to prosecute the coordination of the internal circuit on metadata issues respects. At this phase must undertake the analysis of the methodological documentation and produce a collective opinion, based on:

     - The appreciation of the correct use of concepts and classifications approved by the Statistical Council\(^1\) and the proposals to insert new concepts/classifications on these subsystems (survey based) – issues addressed to the respective subsystem managers: classifications and concepts
     - The appreciation of the correct use of the standard format of the methodological document (also approved by the SC) and the analysis about the coherence of the information – developed by the coordinator
     - The appreciation made on the questionnaire-basis (graphic appearance and the formulation of the questions) – developed by the Questionnaires design expert
     - The appreciation of the survey variables (observation and derived) is not in course yet in a systematic way, only to what indicators respect (dissemination data) – developed by the variables subsystem manager.

  3. **3rd Stage:** The department responsible for the survey updates the methodological document and/or the questionnaire(s) based on the changes admitted proposed by the other Units or justifies its non-agreement. It is obliged to respond to all Units that gave contributions. This documentation - the final (reviewed) version of the methodological document and questionnaire(s) - is then addressed to the Metadata Unit.

\(^1\) The Statistical Council (SC) is the state body that gives guidance and act as coordinator of the National Statistical System (NSS). Amongst other competencies has “To approve technical instruments of statistical coordination, of obligatory appliance in the production of official statistics and to promote its due knowledge, divulgence and use …” - (Law 22/2008 of 13 May).
. **4th Stage**: Facing a new survey or if the changes occurred in a current one refer to methodological aspects, the Metadata Unit on the basis of all the opinions issued by the different units and respective answers, produces a Memo to send to the Board, proposing the survey approval or reporting the arguments that recommend its rejection. The reviewed version of the methodological document and questionnaire are sent ensemble. **Deadline**: 3 weekdays.

If the survey is in neither of the situations signalled above the Metadata Unit acts towards the predicted in the Final degree.

. **5th Stage**: The Board approves or rejects the production of the survey: in case of approval, the methodological document and the questionnaire(s) become final; if it rejects them, the process restarts in order to act on the non-conformities. **Deadline**: 5 weekdays.

. **Final**: As result of the 4th or a successful 5th stage, the Metadata Unit records the questionnaire(s) in the data collection instruments database (subsystem), giving it a registration number and a validity period; and publishes the methodological document (through the appropriate subsystem): on the internal net - Intranet - and on the Official Statistics website (portal). **Deadline**: 5 weekdays.

- **For surveys on the responsibility of delegated entities:**

  . **1st Stage**: The delegated entity prepares the preliminary methodological document and the questionnaire(s) and sends this documentation to INE-PT, to its Methodology and Information Systems Department. Internally, the Metadata Unit becomes in charge of coordinating the request for technical contributions from the units/departments directly involved, such as the Production Department related to the statistical area in focus, the Planning Unit, the Methodology Unit, and is also responsible to ask for in-house contributions derived from the Metadata components/subsystems.

  In the Metadata Unit the methodological documentation manager prosecutes the coordination of the internal circuit - metadata issues. Must undertake the analysis of the methodological documentation based on:

  - The appreciation of the correct use of concepts and classifications approved by the Statistical Council and the proposals to insert new concepts/classifications on these subsystems (survey based) – issues addressed to the respective subsystem managers: classifications and concepts
  - The appreciation of the correct use of the standard format of the methodological document (also approved by the SC) and the analysis about the coherence of the information – developed by the coordinator
  - The appreciation made on the questionnaire-basis (graphic appearance and the formulation of the questions) – developed by the Questionnaires design expert
  - The appreciation of the survey variables (observation and derived) is not in course yet in a systematic way, only to what indicators respect (dissemination data) – developed by the variables subsystem manager.

  On the basis of the opinions given by all units involved (Metadata included) it’s his duty to produce the institution final appraisal. If in this conclusion stage there are no technical objections, the coordinator (also in charge of the data collection subsystem) records the questionnaire(s) in the data collection instruments database, giving it a registration number and a validity period.

  **Deadline**: 10 weekdays;
. **2nd Stage**: The entity responsible for the survey updates the methodological document and/or the questionnaire(s) based on the recommendations emitted in consequence of the technical approval process. This documentation - the final (reviewed) version of the methodological document and questionnaire(s) - is then newly addressed to INE-PT/Metadata Unit.

. **3rd Stage**: The manager of the subsystem publishes the methodological document: on the internal net - Intranet - and on the Official Statistics website (portal).
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 the surveys have a standard format in order to facilitate and increase their usability (Figure 2). This format was approved by the Statistical Council in 2005. Its decision expressed that all surveys concerning the National Statistical System, that is to say, all that represent official statistics should be documented in accordance to this format.

The range of official statistics include the surveys that are INE-PT responsibility as well as the ones that are developed by the delegated entities (state agencies/ministries with which the National Statistical Institute has drawn protocols delegating competencies that allow their own production and dissemination of official statistical data).

This standard/normalised template is subdivided in several groups of information:

<table>
<thead>
<tr>
<th>I. General characterization</th>
<th>Code/ Version / Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIGINE code (information derived from Planning Unit)</td>
<td></td>
</tr>
<tr>
<td>Name/ Survey title</td>
<td></td>
</tr>
<tr>
<td>Statistical activity / Statistical domain</td>
<td></td>
</tr>
<tr>
<td>Purpose (survey)</td>
<td></td>
</tr>
<tr>
<td>Description (summarized)</td>
<td></td>
</tr>
<tr>
<td>Responsible entity</td>
<td></td>
</tr>
<tr>
<td>Relationship with EUROSTAT/ Other entities</td>
<td></td>
</tr>
<tr>
<td>Financing (sources)</td>
<td></td>
</tr>
<tr>
<td>Legal frame (national/communitary legislation/SC decisions)</td>
<td></td>
</tr>
<tr>
<td>Obligatoriness (official statistics – yes or no; Eurostat demands data)</td>
<td></td>
</tr>
<tr>
<td>Type of survey (sample survey or census; analytical/statistical study)</td>
<td></td>
</tr>
<tr>
<td>Type of data source (direct source: survey, administrative acts; non-direct: administrative acts, other surveys)</td>
<td></td>
</tr>
<tr>
<td>Periodicity (annual; infra/supra annual; non periodic)</td>
<td></td>
</tr>
<tr>
<td>Geographic coverage</td>
<td></td>
</tr>
<tr>
<td>Users (list and identification of their particular needs of information)</td>
<td></td>
</tr>
<tr>
<td>Begin/ End date (release date)</td>
<td></td>
</tr>
<tr>
<td>Products (quality standard for disseminating data; type of products by user): Elements requested: Name; Type of product (from table); Frequency (table); Geographical level of data (maximum); Dissemination type (table); Users</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Methodological characterization</th>
<th>Population and Target population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame (physical file containing the target population; origin)</td>
<td></td>
</tr>
<tr>
<td>Sample unit</td>
<td></td>
</tr>
<tr>
<td>Observation unit</td>
<td></td>
</tr>
<tr>
<td>Sample design (if applicable) (sample characteristics; selection methods; dimension; updating methods; software)</td>
<td></td>
</tr>
<tr>
<td>Questionnaire design (methodology; average response time; survey testing)</td>
<td></td>
</tr>
<tr>
<td>Data collection: Direct sources: Reference period for data; Collection period; Expedition date; Initial contact (mean); Collection methods (table); Insistences (nº)/Refusal procedures; Criteria for survey closure; Proxy admittance; Use of incentives to response; Respondents support (means available); Training (interviewers/internal work team) – nº of days/handbook; Data capture (include elements related to: Methods used (table); Coding type (automatic/manual); Software) Administrative acts/Other surveys: Description of the information sources; Purposes; Advantages/Disadvantages regarding its use; Reference period of the data; Target population; Achievement means (electronic or paper)</td>
<td></td>
</tr>
<tr>
<td>Editing (validation rules; methods for data analysis; statistical models)</td>
<td></td>
</tr>
<tr>
<td>Imputation (if applicable)</td>
<td></td>
</tr>
<tr>
<td>Estimation and data achievement (estimators definition and software; explanation on the obtainment of final results)</td>
<td></td>
</tr>
<tr>
<td>Time series (identification of periods and motives)</td>
<td></td>
</tr>
<tr>
<td>III. Concepts</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td>IV. Classifications</td>
<td></td>
</tr>
<tr>
<td>V. Variables</td>
<td>Observation variables</td>
</tr>
<tr>
<td></td>
<td>Indicators (dissemination data)</td>
</tr>
<tr>
<td>VI. Data collection instruments</td>
<td>Respondent entity</td>
</tr>
<tr>
<td></td>
<td>Files</td>
</tr>
<tr>
<td>VII. Abbreviations and acronyms</td>
<td></td>
</tr>
<tr>
<td>VIII. Bibliography</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2 - Standard format of the methodological document

**Methodological Document definition** – product (output) that gathers, in a coherent way, all the metadata generated during the survey design phase.

The *methodological documents subsystem* was implemented in the metadata system in 2004, as a prototype.
Through **the user profile / user interface (general access area)** is possible:

- To consult methodological documents in force in the National Statistical System
- To search methodological documents according to: dissemination theme, survey code, documents’ version, by insert a string pertaining to the survey’s name
- To print the documents’ layout
- To download documents using PDF features.

**The methodological documents subsystem manager (restricted access area)** has several roles, such:

- To ensure that the documents received from the Production Departments comply with the standard format and that the contents of each topic are in accordance with the template
- To check the coherence, harmonisation and completeness of the information
- To prosecute the coordination of the internal circuit on metadata issues in its participation on the technical approval process and assure the deadlines accomplishment
- To coordinate the process of technical approval of surveys between the Units/Departments involved (the Methodology and Information Systems Department is the in-house coordinator)
- To enter the methodological documents into the database. There are two phases of insertion: the first occurs when the methodological document is in its preliminary stage – in the subsystem it will stay with “Interim status”; and the second happens when the document is returned to the Metadata Unit already reviewed by the responsible department and approved by the Board (as a result of the technical approval circuit) – it is then loaded/adjusted in the subsystem and its status is changed to “In force”.
- To assure their publication - for internal consultation (Intranet) and external (Official Statistics website - Portal). There are two phases of insertion: the first occurs when the methodological document is in its preliminary stage – in the subsystem it will stay with “Interim status”; the second happens when the document is returned to the metadata unit already reviewed by the responsible department and approved by the Board (as a result of the technical approval circuit) – it is then loaded/adjusted in the subsystem and its status is changed to “In force”. Only “In force” documents are make available for external consultation.
I.1.2. Data collection instruments subsystem

The data collection instruments subsystem stores and publishes all the questionnaires related to statistical surveys registered by INE-PT (files option is not available, still in preparation). It represents an instrument of reference on data collection. Questionnaires images’ are available as well as the consultation of some items based on their characteristics.

At present time little more than 500 questionnaires are at user disposal, however, since this subsystem was newly added to the metadata system – incremented in 2007 – and by the fact that we are in gradual transition to migrate all the existing data collection instruments, the total would achieve (by now) approximately 3500 questionnaires.

The **main purposes of this subsystem** are:

. To constitute a repository of data collection instruments used in NSS surveys (it goes back to 1961)
. To constitute a management tool for collection instruments

There are basically two types of statistical data collection instruments:

. Questionnaires
. Files (not implemented by now)

**Data Collection Instrument definition** - Instrument used for data collection and for which an identification can be specified by means of an attribute.

The register of a data collection instrument is the final act of the technical approval process; it stands for a label that certifies and assures the survey overall quality.

This certification – the registration of a particular data collection instrument - always comprehends two aspects:

. The assignment of a registration number (numeric four digits code, automatically addressed by the subsystem)
. The assignment of a certain validity period (defined by the survey responsible, in most cases a one-period) – surpassed this time period the data collection instrument is no longer in force.

**To register a data collection instrument** two situations are in balance, only one may occur based on the real case:

. The record is inserted in the database subsystem as a **new** register, because:
  - The collection instrument is related to a new survey
  - In a routine survey there have been changes to the content of a collection instrument. They may be due to:
    - Inclusion or exclusion of variables
    - Changes to questions (while modifying the essence of the object of study)
    - Change on the name of the survey

  Implies a **substitution** performance
The record already exists in the database, maintains its characteristics and only requires the revalidation of the term period – a new time period of validity is given. 

Implies a prorogation performance

This subsystem makes it possible to the user profile / user interface (general access area):

- To consult questionnaires information (characterization elements)
- To consult the history of different collection instruments
- To view questionnaires images’ and print their layouts
- To identify and accede the methodological documents associated (PDF file)
- To find out what variables they observe (with direct link to this subsystem)
- To export the result of a search operation

The Detail information includes a number of items concerning each questionnaire:

- Status (in force or not in force)
- Registration number
- Name (data collection instrument)
- Validity
- Geographic coverage
- Periodicity
- Type of instrument (questionnaire or file)
- Questionnaire (image in PDF format)
- Data collection method
- Respondent entity
- Statistical observation unit
- Statistical area
- Responsible entity
- Survey: access and consultation of the Methodological Document related to the instrument
- Historic information (identifies eventual Prorogations and Substitution cases: giving evidence of the substituted and the respective substitute instrument)
- Variables

All these components represent modes of search that can be chosen per se or combined, once the user opts by:

- Search (search by a single element) or Advanced search (search by one or more elements using logic operators in a combination of conditions that can be operated as an cumulative (using the “and” operator) or an exclusive function (using “or”)

The results are presented in a table form; the number of rows shown respect to the number of data collection instruments found that meet the selected criteria. From that list the user may accede to any singular questionnaire, addressing the Detail information view. This scope can be exported as an output – as a CSV file.
The administration profile / subsystem manager interface (restricted access area) comprehends different functions:

- Registration of data collection instruments in the database (questionnaires so far) and respective allocation of validity periods

- Execution of management procedures involving database information: insertion/deletion/adjustments; acts on prorogation/substitution performances; introduces and updates decoding tables

- Control and test the migration of data

- Insertion of translated names (data collection instruments) – English/French

- Management of respondent entities (generation/elimination/adjustments)

- Assurance of users and profiles management (insertion/deletion)

- Interaction with the IT technicians involved in the implementation/maintenance of the subsystem

- Production of reports (information extracted from the subsystem, pursuing: responses to particular needs, regular checks on quality information assessment (coherence, accuracy, relation to elements included in the respective methodological document), anticipation of requests in routine surveys - prorogations).

Bibliography:

- INE-PT/Methodology and Information Systems Department/Systems and Metadata Unit, “Data Collection Instruments subsystem – Conceptual analysis”, June 2008 (reviewed version)

- INE-PT/Methodology and Information Systems Department/Systems and Metadata Unit, “Methodological Document Template – Surveys documentation” (DOCT/1634/CSE/PCD), May 2005


- Morgado, Isabel and Isfan, Monica (INE-PT/Methodology and Information Systems Department/Systems and Metadata Unit), “Statistics Portugal Case study”, (METIS Common Metadata Framework), April 2008