4. Statistical Metadata Systems (United Nations Industrial Development Organization (UNIDO))

back to case study home page

4.1 Metadata system(s)

18. The metadata system is a part of the Integrated Statistical Development Environment, provides end-to-end metadata services throughout the statistical production process and was developed in the context of the migration from Mainframe to a Client Server Platform. Figure 2 presents the overall structure of ISDE and its relation to the statistical production life cycle. The client part of the system is presented to the user as a desktop application, the **ISDE shell**that serves as a container for the rest client/side applications. These applications are described briefly below.

- **ADMIN** - provides administrative services, like user and authorisation management, logging and auditing of the system, backup and restore management;
- **Nomenclature Explorer** is the tool for maintenance of the core definitional metadata, which is not related to particular data items but rather serve for defining the structure of the data and metadata. These first two applications are outside of the life cycle. ***Figure 3*** and ***Figure 4*** show examples of the ISDE shell;
- **Questionnaire** is the application for management of the pre-filling and distributing of the questionnaires to the member countries (i.e. used in the *Initialisation* phase);
- **Data Wizard** is the main data and metadata maintenance tool used in the Data Collection and Transformation phases of the life cycle. It provides services for:
  i. Reading in the data and metadata from the returned back Excel questionnaire
  ii. Initial validation of the read in data and storing in the database (at stage 1)
  iii. Maintenance of the metadata iv. Screening
  v. Aggregation and further data validations and transformations
- **Presentation Wizard** is mainly a visualization tool which can be used in the Dissemination phase for answering ad hock requests, but because of its versatile functionality it finds a wide usage also in the *Data Transformation*phase
- **Publication applications** - these are the applications used in the Dissemination phase for generating the different publication products
  i. Yearbook - this a complex set of applications for production of the Industrial Statistics yearbook including aggregation, layout, PDF file generation according to pre-defined templates and other tools. The final result is a publication ready PDF file of about 700 pages;
  ii. INDSTAT CD - used to produce the INDSTAT type of CD products;
  iii. IDSB CD - used to produce the INDSTAT type of CD products;
  iv. WEB - used to generate the necessary data and metadata for updating the WEB dissemination database (this database is outside of the ISDE system, managed by the computer section);
- **Other applications** - in this category are included any other applications used in the process, like SAS, R, tools for compilation of Production index numbers and National Accounts data (which are outside of the scope of this document) and others.

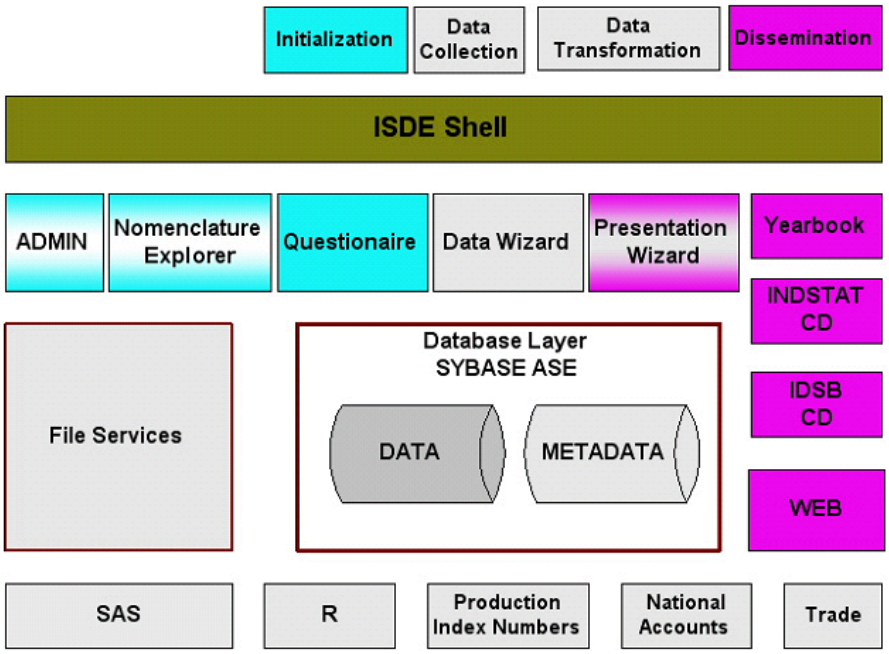

Figure 2: Overall structure of the Integrated Statistical Development Environment (ISDE) and its relation to the statistical production life cycle

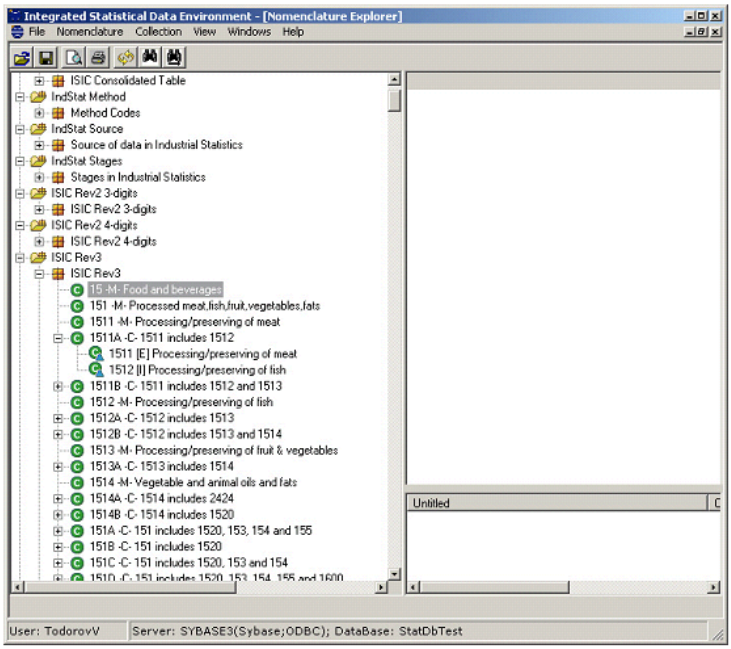

Figure 3: The tool for maintenance of the core definitional metadata - Nomenclature Explorer- running in the ISDE shell

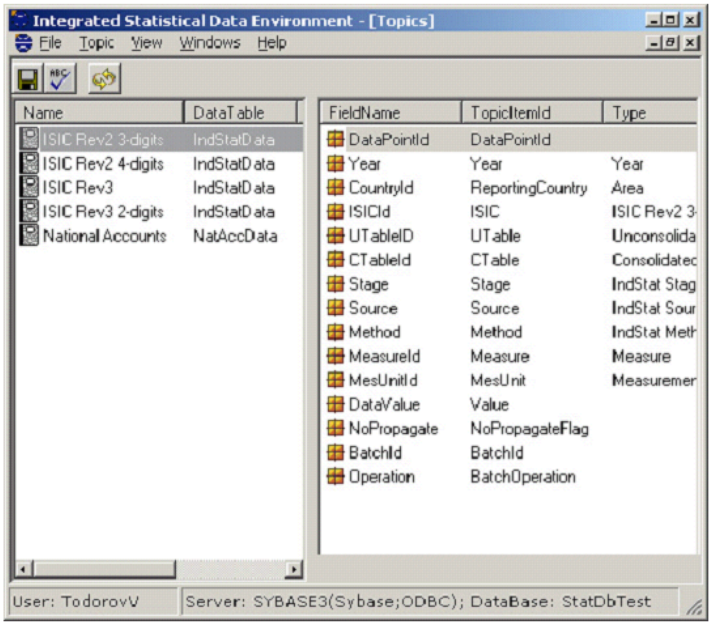

Figure 4: Defining topics, i.e. logical databases in ISDE- the ADMIN module running in the ISDE shell

19. As already mentioned the ISDE was developed in the context of migration from Mainframe to a Client/Server platform. For the migration a stepwise approach was chosen because of the following reasons:

- The project was not urgent, since the discontinuation of the mainframe was postponed because of other important services still running on it
- The software test and sustaining of the created system has to be done in-house
- Only limited resources were available
- The staff was very willing to participate in the project
- The goal was not only to migrate the system but rather to develop a completely new one and the requirements were not yet completely specified (because of the limited resources)
- A key requirement was that the established UNIDO data services must not be disrupted

20. The first step was a rigorous analysis of the existing system and development of a data model which was as generic as possible in order to be able to accommodate any changes. Based on this model a loader application was developed which allowed in any moment to synchronize the data in mainframe and in the Sybase database of the new Client/Server system. The development of the new metadata subsystem was initiated by implementing a tool for maintenance of the definitional metadata [2005-2006]. Thus a kind of proof of concept was successfully completed.

21. A capture/maintenance tool for reference metadata was developed and the description/methodological metadata, which existed so far in the form of Word documents or Excel worksheets, were entered into the system. The mainframe footnote database (data-item level metadata) was imported too. Thus the complete process of maintenance of the available metadata was migrated to the Client/Server platform

22. In the next step the data dissemination applications were developed which allowed to produce the recurrent statistical publications/products from the mainframe system and from the Client/Server platform in parallel which was an ideal acceptance test for the new applications by just comparing the results [Q4-2006 - Q4-2007].

23. As an example of the migration-to-new development relation can be noticed that while the International Yearbook of Industrial Statistics was produced from the main frame as a cameraready line printer output which was glued together with many MS Word and MS Excel documents, the output of the Client/Server system was an automatically generated page numbered PDF file of about 700 pages.

24. In the third step the pre-filling of the questionnaire was implemented using the new Client/Server data- and metadata-base [Q1-2007]. The data capturing as well as the data maintenance tools were developed and are now in the phase of final testing. The questionnaires, which are expected to start arriving in June, will be entered only in the Client/Serve system. This will be the ultimate decoupling of the new system from the mainframe.

44. The overall structure of the Integrated Statistical Development Environment is presented in Figure 2. The system utilizes a 3-tier architecture build on .Net technology. The data and metadata are stored in centralized database, and the user interacts with the system through the ISDE shell which is a desktop application serving as a container for the other ISDE applications. The commonality of the system is achieved through

using shareable component libraries.

45. The development of the entire Integrated Statistical Development Environment has been carried out inhouse, taking international standards (ISO/IEC 11179) into consideration.

Application/Tool	Description
Sybase ASE 12.5	"Adaptive Server Enterprise" - the relational database management software manufactured and sold by Sybase, Inc. There are two separate databases running on the Sybase server – a test and production one. Another couple of test/production databases is used for web publications, but it is completely outside of the statistics unit.
Erwin 4.1	Data modelling and database maintenance
SQL Programmer 12:0	Database maintenance
.NET Framework 2.0	Most of the applications are written in C#
Crystal reports	A general purpose reporting tool, used for the production of the Industrial Statistics Yearbook as well as other publications. The version bundled with MS Visual Studio was used
MS Visuals Studio 2005	The main tool used for the development of the Client/Server libraries and applications
XML Spy 2.0	Advanced XML editor
ISDE File services	Used for interactions between ISDE and other tools like SAS and R. This is a shared network drive onto which the ISDE users have access.
SAS	Used for processing the National accounts data, the Production Index numbers as well as for serving any ad hock requests for data
R	Currently used only for very specialized tasks, very high graphical potentials
VB 6.0	There are several legacy tools written in VB 6.0 which are not yet ported to the Microsoft .Net Framework (migration pending)

Table 7: List of tools used for the development of the system

Database layer

46. The database consists of two identical but physically separated databases - a test and production databases - running on Sybase ASE RDBMS under Linux. A sample of the data model is shown in Figure 14 and the complete data model is presented in an attached Erwin diagram.

47. The access to data and metadata from the client applications is performed through component libraries. These would allow replacing for example the Sybase database by an MS SQL Server or Oracle without any modification of the applications.

Component libraries

48. The object oriented component libraries are developed also in C# and are used to unify many common tasks like database access, file access, printing, access to common data structures, etc.

Client applications

49. The client applications are developed using MS Visual studio in C#. They connect to the database and interact with each other using component libraries developed also in C#.

Other tools

50. Table 1 lists some other tools integrated in the ISDE system.

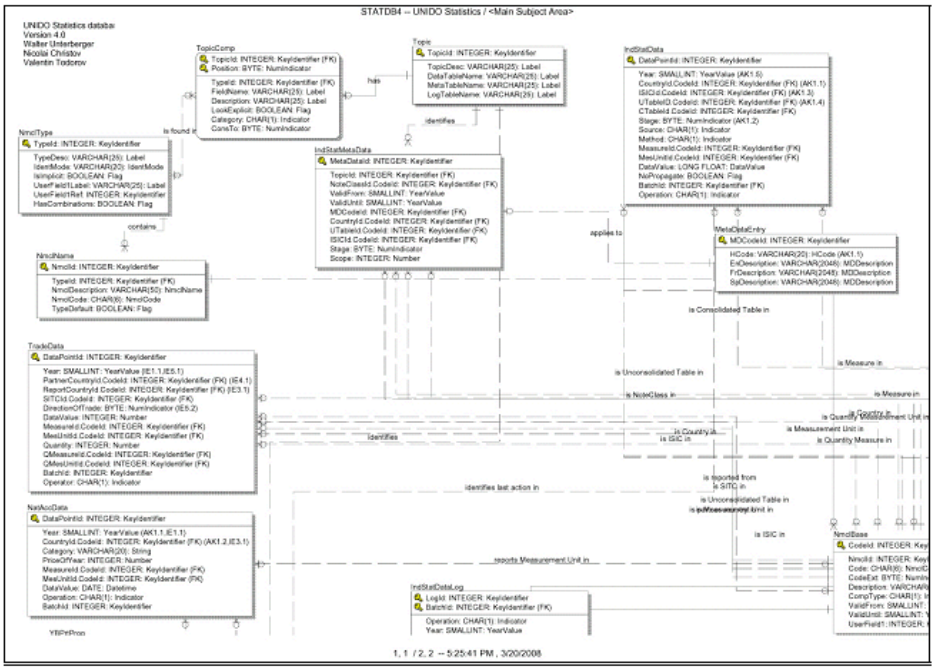

Figure 14: Part of the data model of the UNIDO Industrial Statistics data- and metadata-base. The complete data model is given in the attachements