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

**UNECE Workshop on the Common Metadata Framework**
(Luxembourg, 9-11 April 2008)

**CASE STUDY – UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANISATION (UNIDO) [1]**

# 1. INTRODUCTION

## Organization Details

| | |
|---|---|
| Organization Name: | United Nations Industrial Development Organization (UNIDO) |
| | Research and Statistics Branch |
| Contact: | Valentin Todorov |
| E-mail: | v.todorov@unido.org |

1.  UNIDO was set up in 1966 and became a specialized agency of the United Nations in 1985. As part of the United Nations common system, UNIDO has responsibility for promoting industrialization throughout the developing world, in cooperation with its 172 Member States. Its headquarters are in Vienna, and it is represented in 35 developing countries. This representation and a number of specialised field offices, for investment and technology promotion and other specific aspects of its work, give UNIDO an active presence in the field.

2.  The Service Module "Industrial Governance and Statistics" assists developing countries and economies in transition to **monitor, benchmark and analyse their industrial performance and capabilities**, and on that basis to **formulate, implement and monitor strategies, policies and programmes** to improve the contribution of industry to productivity growth and the achievement of the UN Millennium Development Goals (MDGs).

3.  Building capabilities in industrial statistics: UNIDO provides technical assistance to introduce best practice methodologies and software systems to monitor and assess productivity performance and use it as a guide for policy-making; and enhance the quality and consistency of the industrial statistics databases so as to provide meaningful inputs for assessing productivity and industrial performance.

---

[1] Prepared by Valentin Todorov and Shyam Upadhyaya

4.  Most developing countries lack the capacity to systematically gauge productivity performance in such a way as to be able to use it as an effective tool for policy and decision-making in general. Through this service module UNIDO aids developing countries and countries with economies in transition to build capacities in this field by providing technical assistance to:

    - Introduce best practice methodologies and software systems to monitor and assess productivity performance and use it as a guide for policy-making; and
    - Enhance the quality and consistency of the industrial statistics databases so as to provide meaningful inputs for assessing productivity and industrial performance.
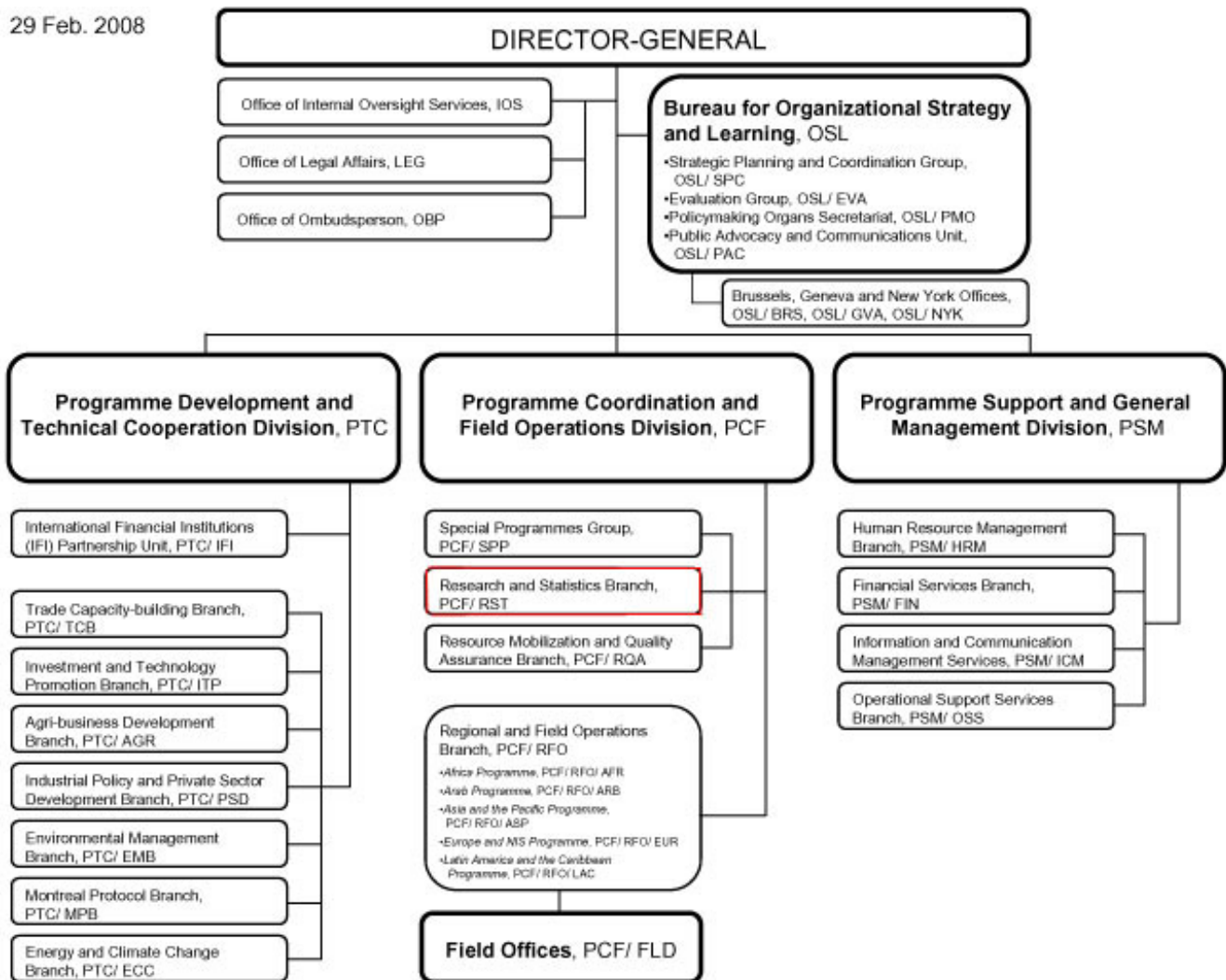


*Figure 1: UNIDO Organization Chart: Statistical activities are carried out by the Statistical Unit of the Research and Statistics Branch PCF/RST/STA*

5.  Statistical activities of UNIDO are carried out by the Research and Statistics Branch.

## Overall strategy and metadata management principles:

6. The conceptual development of the UNIDO metadata subsystem was initiated in 1999 with the aim of automation of information production (data and metadata) using latest management technology. Having in mind the inherent structural complexity of the data bodies involved, only a comprehensive metadata-based system re-design approach has been considered promising at all. Thus, the project favoured an **integrated data and data documentation (metadata) framework** emphasizing that, while allowing scrutiny of data documentation (statistical metadata) both individually and jointly with statistical data, any statistical data access always entails the retrieval of associated metadata without demanding specific inquiry measures or actions. This way a rather tight interrelation of data and metadata is both enforced and assured by purely technical means. However, as its major precondition, this principle presupposes a homogenous representation of all pieces of data documentation in order to enable uniform data and documentation access procedures.

7. Moreover, as a change in data representation must not disrupt established UNIDO data services, **a smooth migration policy** is called for, leaving interface requirements of downstream systems and data usage almost untouched. Implying such a great effort to UNIDO, an expected side-benefit of re-designing the INDSTAT system is its potential applicability to further operational data management areas in need of refashion.

8. The concrete design and implementation of this subsystem was realized as a part of an integrated data and metadata system under the name **Integrated Statistical Development Environment (ISDE)**. The development of ISDE was performed in a stepwise manner in the context of a migration project of the complete UNIDO statistical databases from an IBM mainframe to a client/server platform. Further in this document will be given details about the migration project itself, its current status and its relation to the newly developed statistical applications and ICT infrastructure.

9. The basic metadata management principles are backed by the **UNIDO Quality Assurance Framework** which is targeted to ensure that the statistical activities of UNIDO are relevant and the data compiled and disseminated are accurate, complete within the defined scope and coverage, timely, comparable in terms of internationally recommended methods and classification standards and internally coherent to variables included in the datasets. While these generally accepted, broad dimensions of quality of statistical data may be defined in each NSO's own quality assurance framework, UNIDO makes maximum effort that data produced from the statistical operation undertaken with the UNIDO technical cooperation are accurate, internationally comparable and coherent. UNIDO has been a forerunner among international organizations in using a Statistical Metadata System as a tool for observing and evaluation the quality of statistical data, especially for completeness and cross-country comparability. For further details see Yamada (2004) and Upadhyaya (2008).

10. Following the **International recommendations for Industrial Statistics,** the development of metadata is given a high priority and their dissemination is considered an integral part of dissemination of industrial statistics. Moreover, it is recommended that in consideration of the integrated approach to compilation of economic statistics development of a coherent system and a structured approach to metadata across all areas of economic statistics be adopted, focusing on improving their quantity and coverage. Further, the dissemination of statistical data and metadata using web technology and SDMX standards is recommended as a way to reduce the international reporting burden (the **Statistical Data and Metadata Exchange (SDMX)** technical standards and content-oriented guidelines provide common formats and nomenclatures for exchange and sharing of statistical data and metadata using modern technology).

11. An essential requirement was that the metadata is available in three languages (English, French and Spanish). This allows to pre-fill each questionnaire in the preferred language for the country and than to process it accordingly.

12. The integrated system is based on a formal framework, described in detail in Froeschl et al. (2002), Froeschl and Yamada (2000). The proposed information system architecture comprises two cubes, one for statistical data and another for the metadata interrelated by a set of shared dimensions.

## 2. THE STATISTICAL METADATA SYSTEMS AND THE STATISTICAL CYCLE

13. Statistical activity of UNIDO started with establishment of Industrial Statistics database in 1977 to meet the internal needs of the organization for an accurate assessment of structure and growth of industrial sector. In terms of the external data sources UNIDO was dependent to the United Nations Statistics Division, which used to collect the data from national statistical offices. In 1993 UN Statistical Commission at its twenty-seventh session granted mandate to UNIDO in collaboration with OECD for collection, maintenance and dissemination of worldwide key industrial statistics.

14. Currently UNIDO maintains **industrial statistics databases** or INDSTAT databases which are regularly updated with the data collected from National Statistical Offices (NSOs) and OECD (for OECD member countries). UNIDO also collects national account related data from National Accounts Main Aggregate Database of UNSD, the World Development Indicators of the World Bank and other secondary sources. Further in this paper mainly the INDSTAT databases will be considered and other data sources will be mentioned only as needed as complimentary data for particular statistical dissemination products. Thus the statistical data production life cycle model of UNIDO (we do not talk about statistical survey life cycle) consists of collection, development, maintenance and dissemination of worldwide key industrial statistics at detailed sub-sectoral levels of the manufacturing sector of industrial production in individual countries.

15. The process of statistical production in UNIDO has an annual cycle and can be summarized as follows. Elsewhere in this document each phase of the lifecycle is described in more detail.

    o **Initialisation**: [Q1] Pre-filling of the out-going UNIDO General Industrial Statistics Questionnaire with previously reported statistical data and metadata for their possible revision by the NSO. The questionnaire is created in Excel format in one of the three languages (English, French or Spanish) appropriate for the particular country. The pre-filling is automated using the available data and metadata;

    o **Data Collection (NSO)**: [Q2-Q3] The completed and returned to UNIDO by the NSO questionnaires (excel format, rarely hard copy) are entered into the system and are ready for further validation and processing;

    o **Data Collection (OECD):** [Q3] Data for OECD member countries, collected through joint OECD/UNIDO questionnaire and transmitted to UNIDO (excel format) are entered into the system and are ready for further validation and processing;

    o **Transformation/Processing**: [Q3-Q4] The data collected from the primary sources are further transformed to a ready-to use data set. The data transformation is done in five stages, which not only constitute an operational framework for UNIDO statisticians, but also provides additional description of statistics (generated metadata which is attributed to each data item) to users.

o **Dissemination:** [Q4-Q1] After undergoing the complete processing phase the incoming and generated data and metadata are stored in the databases and the databases and can be used for production of the recurrent statistical publications: International Yearbook of Industrial Statistics, INDSTAT and IDSB CD products, Web Country Statistics.

16. Similarly to the survey life cycle defined by the METIS group, at the beginning can be an optional 'Need' phase in which the necessity of changes in the questionnaire and processing is analysed. If changes are required, the respective 'Develop and Design' phase is in order. For example this year the metadata part of the questionnaire was evaluated and modified, which included translation into the three supported languages (English, French and Spanish).

17. The following *Table 1* shows the mapping between METIS survey live cycle phases and the phases in the UNIDO production process. Essential difference is that there are no explicit preparation phases like 'Need', 'Design and develop', but always at the beginning of the statistical production cycle the current status is analysed and if necessary the questionnaire as well as the process are updated. Further, there is no 'Archive' phase, since as soon as the data are processed completely they are stored in the UNIDO statistical databases and there is no need of special archiving procedure.

| METIS | UNIDO |
|---|---|
| Need | Need [optional] |
| Develop and design | Develop and design [optional] |
| Build | Initialisation: pre-fill and distribute questionnaires |
| Collect | Data Collection |
| Process | Transformation/Processing |
| Analyse | Analysis |
| Disseminate | Dissemination |
| Archive | - |
| Evaluate | Evaluation |

*Table 1: Mapping of the UNIDO cycle phases to these developed by the METIS group*

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.

o **ADMIN** – provides administrative services, like user and authorisation management, logging and auditing of the system, backup and restore management;

o **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;

o **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);

- o **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

- o **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

- o **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);

- o **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.
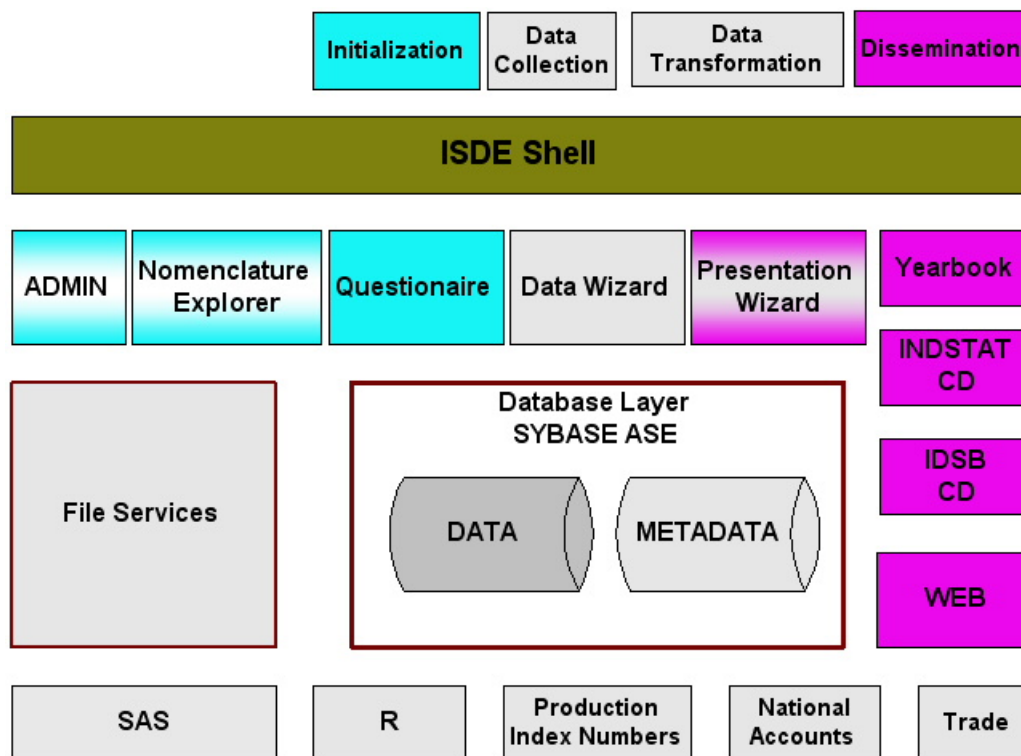


*Figure 2: Overall structure of the Integrated Statistical Development Environment (ISDE) and its relation to the statistical production life cycle*
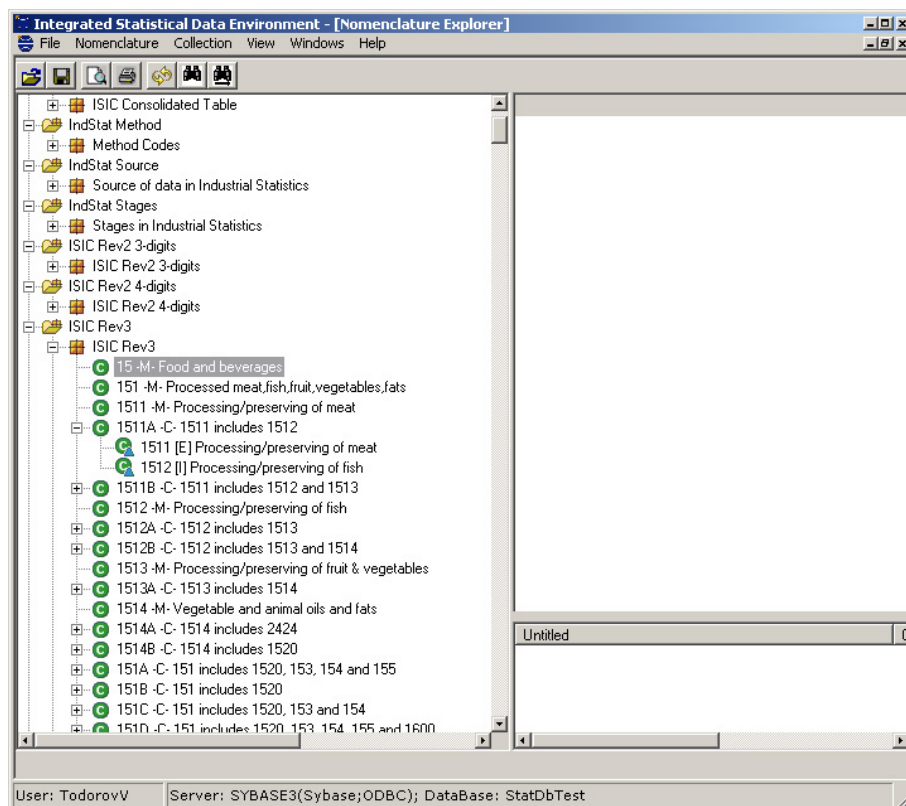
*Figure 3: The tool for maintenance of the core definitional metadata - Nomenclature Explorer- running in the ISDE shell*
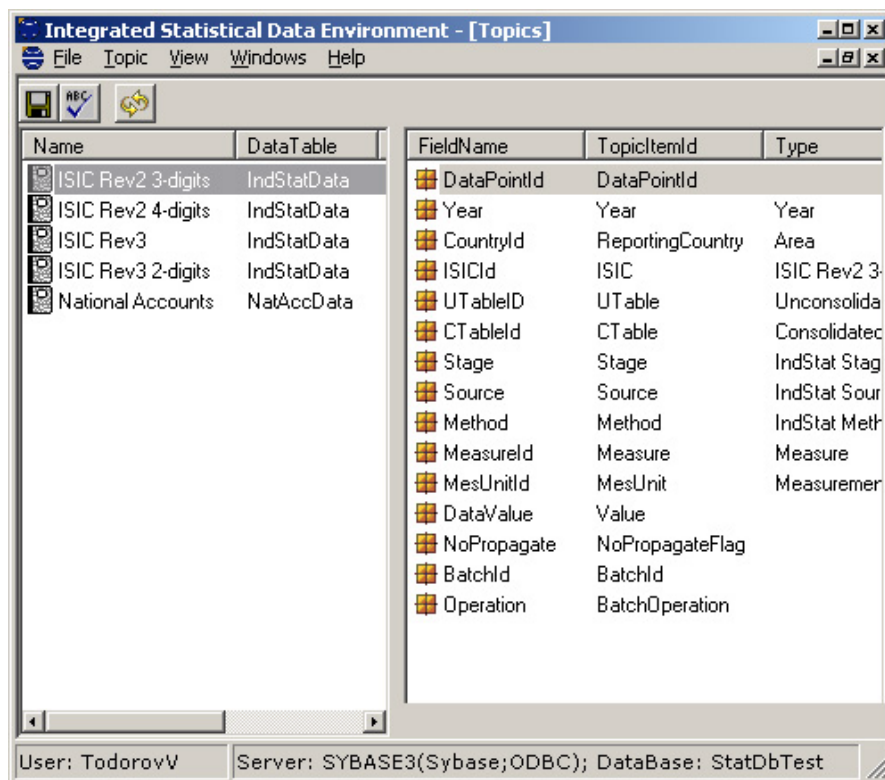


*Figure 4: Defining topics, i.e. logical databases in ISDE- the ADMIN module running in the ISDE shell*

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:
    o The project was not urgent, since the discontinuation of the mainframe was postponed because of other important services still running on it
    o The software test and sustaining of the created system has to be done in-house
    o Only limited resources were available
    o The staff was very willing to participate in the project
    o 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)
    o 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 camera-ready 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.

# 3. STATISTICAL METADATA IN EACH PHASE OF THE STATISTICAL CYCLE

25. The metadata is classified according to their usage and their role in the statistical production process. The main types of metadata according to this criteria are as follows:

    o **Definitional metadata** – The definitional metadata refer to metadata that act as identifiers and descriptors of the data. They are prior to the data, are created and maintained

independently from the data and are used to define the data structure. Examples of definitional metadata are country names and codes, currency names and codes and their relation to the countries, definitions of the indicators, classifications like ISIC Rev. 2, ISIC Rev. 3, etc. Through these core data are defined also some basic metadata elements like metadata classes, stages, sources and methods, etc. Historically this metadata type was the first to be established (ported from the Mainframe, re-factored and formalized) in ISDE. The definitional data are maintained by the statistical staff using the tool Nomenclature Explorer (NE) following strictly the user authorisation and ownership.

o **Implicit metadata** – The implicit metadata are a special class of metadata arising throughout the specific usage of other metadata. Typical example are the ISIC combinations. For example several industry categories can be combined and reported together by a given country for a given indicator and years. In the questionnaire returned by the NSOs such a combination is expressed in the following way (see – *Figure 6*):

| ... | | |
|------|------------------------------------------|-----------|
| 1511 | Processing/preserving of meat | 1234 **a/** |
| 1512 | Processing/preserving of fish | ... **a/** |
| 1513 | Processing/preserving of fruit & vegetables | ... **a/** |
| ... | | |

```
REMARKS:
a/ 1511 includes 1512 and 1513
```

The codes 1511, 1512 and 1513 are combined and reported as a single number '1234'. The combined industries are linked by the footnote **a/**. This is resolved by the system as a dummy ISIC code 1511A defined as "1511 includes 1512 and 1513" which is used throughout the production process and appears accordingly in the publications as well as in the pre-filled Questionnaire.

In a similar way can be solved other country specific classification discrepancies like industry codes at 3-digit level that **exclude** one or more specific 4-digit industry codes. The implicit metadata can be used also for defining of synonyms – for example '040' is the country code of Austria and this is the same as, i.e. **substituted by** the ISIC code 'AUT'. Or for specifying of aggregation e.g. the aggregation code 'EU' is composed by the codes of the single countries. The keywords **substitute**, **included**, **excluded** used in the above described context are called operators.

o **Operational Metadata** – The operational metadata are generated by the process of data transformation and attributed to the respective data items. As described in the presentation of the Data Transformation phase, each data item is stored in the database with a stage indicator reflecting its credibility. Also the transformation process generates "Source" and "Methods" metadata, describing the source of the data item and methods applied for its generation.

o **System metadata** – these metadata are used to drive automated processing throughout the phases of the life cycle. These can be layout definitions for the yearbook (for each country, for each edition of the yearbook) as well as country lists, etc., used in the automatic generation of the PDF output; Installation and packaging lists, directories, templates, etc. for creation of the CD product. These metadata are specific for the application where they are used and do not relate to the data, therefore, although stored in the centralized

repository, are maintained by each application separately and are called "Properties" of the respective process, i.e. Yearbook properties, Questionnaire properties, etc.

- o **Descriptive and Methodological metadata** – these form the main bulk of metadata. They are received from the primary data reporters, using the UNIDO Questionnaire and than are further processed together with the data. During this processing additional metadata can be added by the UNIDO statistical staff. Descriptive or methodological metadata can be attached to all possible levels ranging from the complete data set down to individual data items. This is done by assigning to the metadata same dimensions as those of the data.

26. In the rest of this section is described each single phase of the statistical production life cycle and for each phase the metadata used or created is specified.

## Initialisation

27. The main output of this phase is the pre-filling of the out-going UNIDO General Industrial Statistics Questionnaire with previously reported statistical data and metadata for their possible revision by the NSO. The questionnaire is created in Excel format in one of the three languages (English, French or Spanish) appropriate for the particular country. The pre-filling is automated using the available data and metadata.



*Figure 5: Metadata in the Preparation Phase: the pre-filled Excel questionnaire contains one worksheet with metadata as well as may contain metadata to each data item.*

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| **Metadata used** | | | | | |
| Definitional metadata | These are metadata serving as identifiers and descriptors of the data | Country codes and names, Country groups, indicators, classifications: ISIC Rev 2, ISIC Rev 3, currency names and codes and their relation to the countries | Defined and maintained by the subject experts using the tool Nomenclature Explorer | | |
| Implicit metadata | These are metadata arising from special application of other metadata | Typical example are the ISIC combinations | Created by NSO, but can be generated also during the processing phase | | |
| Operational metadata | are generated throughout the process of data transformation. Used to select the data for pre-filling of the questionnaire – only Stage 1 data | Stages, sources, methods | Generated during the processing phase and attributed to each particular data item | | |
| Systems metadata | drive automated processing | Template for the questionnaire, Language and ISIC revision used by the respective country, measurement units, etc. | Created for each country by the responsible for the country staff member | | |
| Descriptive and Methodological metadata | Describe the deviations from the international standards and the methodology used to acquire the data | Supplier of information, Major deviations from ISIC, reference period, Concepts and definitions of variables, etc. | Created by NSO, but can be re-described/rearranged by UNIDO in order to comply with the international standards | Increase the relevancy of the data and add to the interpretability of the data | |
| | | | | | |
| **Metadata produced (output)** | | | | | |
| None | | | | | |
| | | | | | |

*Table 2: Metadata used (input) and produced (output) in the Initialisation phase*

**Data Collection**

28. After receiving back the completed questionnaires, they are entered in the system for validation and further processing. The excel file is read automatically in and the user has range of tools for validation, analysis, correcting etc. – see *Figure 7*. During the processing of the particular questionnaire with all data and metadata included it can be stored in the interim storage in XML format. The metadata can be edited or new can be entered – see *Figure 8* and *Figure 9*.

29. The preparation of appropriate statistical metadata in support of the INDSTAT databases requires concrete and well-documented metadata inputs from the primary data compilers. Thus, UNIDO requests NSOs to provide, together with available statistical data, such descriptive information through its industrial statistics country questionnaire. The key items for which the organization needs to obtain metadata include:
    o Name of the supplier of the statistical data (i.e. reporting agency),
    o Basic source of data (e.g., annual industry survey),
    o Data reporting system (major deviations from ISIC),
    o Reference period (e.g. calendar year),
    o Reference unit (type of statistical unit)
        i. Establishment
        ii. Enterprise
        iii. Other
    o Scope of the annual survey (type of reference units covered) – information on coverage and the cut-off size,
    o Employed method of data collection,
    o Employed method of enumeration (direct interview, mail or web-surveys),
    o Response rate,
    o Treatment of non-response,
    o Concepts and definitions of the variables on which data are reported (details about each indicator),
    o Related national statistical publications and

30. The provided metadata are sometimes not described from the viewpoint of international comparability but rather from the viewpoint of national standards. In such cases the UNIDO statistical staff re-describes/rearranges the provided metadata into explicit information for the deviation from the international standard. This is often a difficult task and requires additional meta-information from the concerned NSO.

31. Additionally to each data item in the questionnaire can be attached one or more metadata items (footnotes in the older UNIDO terminology), like "Missing because of confidentiality reasons" or combinations of ISIC codes like "1511 includes 1512", etc – see *Figure 5*.

32. The metadata that are provided by NSOs often do not explicitly indicate deviations from international standards. In such cases, UNIDO attempts to r-describe/re-arrange the provided metadata into explicit information concerning the deviations from the international standards. This is often a difficult task and requires additional clarifications from the concerned NSO.

33. Data for OECD member countries, collected through joint OECD/UNIDO questionnaire and transmitted to UNIDO (excel format) are entered into the system in a similar way and are ready for further validation and processing. These questionnaires do not contain metadata, which is extracted from other OECD publications - OECD (2003) Industrial Structure Statistics, Volume 1, Core Data

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| **Metadata used** | | | | | |
| Definitional metadata | These are metadata serving as identifiers and descriptors of the data | See *Table 2* | See *Table 2* | | |
| Implicit metadata | These are metadata arising from special application of other metadata | Typical example are the ISIC combinations | Created by NSO, but can be generated also during the collection and processing phase | Contribute to the cross country comparability | The automated combination resolution while importing the data and metadata from the Excel questionnaire can be quite difficult |
| Descriptive and Methodological metadata | Describe the deviations from the international standards and the methodology used to acquire the data | Supplier of information, Major deviations from ISIC, reference period, Concepts and definitions of variables, etc. | Created by NSO, but can be re-described/rearranged by UNIDO in order to comply with the international standards | Increase the relevancy of the data and add to the interpretability of the data | |
| | | | | | |
| **Metadata produced (output)** | | | | | |
| Operational metadata | are generated throughout the process of data collection and transformation | Stages – the collected data is stored at stage 1, the source is set to NSO | Generated during the collection and processing phase and attributed to each particular data item | | |
| | | | | | |

*Table 3: Metadata used (input) and produced (output) in the Data Collection phase*

# Figure 6

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | UNIDO: General Industrial Statistics Questionnaire | | | | | | | 2005 Edition |
| 2 | Country: ALBANIA | | | UN Code: 008 | | | | Table 05 |
| 3 | | | | Wages and salaries paid to employees | | | | |
| 4 | | | | (thousands of Albanian Leks) | | | | |
| 5 | ISIC Rev. 3 | INDUSTRY | Note | 2001 | 2002 | 2003 | 2004 | 2005 |
| 153 | 3694 | Games and toys | | … | … | … | … | … |
| 154 | 3699 | Other manufacturing n.e.c. | | … | … | … | … | … |
| 155 | 3710 | Recycling of metal waste and scrap | | … | … | … | … | 15491q/ |
| 156 | 3720 | Recycling of non-metal waste and scrap | | … | … | … | … | …q/ |
| 157 | D | Total manufacturing | | 5173538p/ | 5988617p/ | 8703238p/ | 9004800p/ | 9228711p/ |
| 158 | REMARKS | | | | | | | |
| 159 | a/ 1511 includes 1512. | | | | | | | |
| 160 | b/ 1531 includes 1532. | | | | | | | |
| 161 | c/ 154 includes 1533. | | | | | | | |
| 162 | d/ 171 includes 172, 1730, 1810 and 1820. | | | | | | | |
| 163 | e/ 191 includes 1920. | | | | | | | |
| 164 | f/ 2010 includes 202. | | | | | | | |
| 165 | g/ 210 includes 221, 222 and 2230. | | | | | | | |
| 166 | h/ 2310 includes 2320 and 2330. | | | | | | | |
| 167 | i/ 241 includes 242, 2430, 251 and 2520. | | | | | | | |
| 168 | j/ 2610 includes 269. | | | | | | | |
| 169 | k/ 2710 includes 2720, 273, 281, 289, 291, 292 and 2930. | | | | | | | |
| 170 | m/ 3000 includes 3110, 3120, 3130, 3140, 3150 and 3190. | | | | | | | |
| 171 | n/ 3610 includes 369. | | | | | | | |
| 172 | p/ Sum of available data. | | | | | | | |
| 173 | o/ 241 includes 2310, 2320, 2330, 242, 2430, 251 and 2520 | | | | | | | |
| 174 | q/ 3710 includes 3720 | | | | | | | |
| 175 | | | | | | | | |

*Figure 6: Metadata in the Collection Phase: the returned Excel questionnaire may contain metadata to each data item. These metadata items are read and processed automatically by the system*

# Figure 7

Data Wizard [todorovv@StatDb]
File  Data Sheet  Meta Data  Tools  View  Window  Help

[100] Bulgaria/2005 --- [04] Number of employees [SCALE=Units] [C=04]

100 -M- Bulgaria
  2005
    02 -M- Number of
    03 -M- Number of
    04 -M- Number of
    05 -M- Wages and
    13 -M- Output in p
    19 -M- Value adde
    21 -M- Gross fixed
    31 -M- Number of
  2004
  2003
  2002
  2001

| ISIC Rev3 | INDUSTRY | ISIC1 | Stage1 | ISIC2 | Stage2 | ISIC3 | Stage3 | ISIC |
|---|---|---|---|---|---|---|---|---|
| 15 | Food and beverages | 15 | | 15 | | 15 | | 15 |
| 151 | Processed meat,fish,fruit,vegetables,fats | 151 | 27411 | 151 | 27411 | 151 | 27411 | 151 |
| 1511 | Processing/preserving of meat | 1511 | 15278 | 1511 | 15278 | 1511 | 15278 | 1511 |
| 1512 | Processing/preserving of fish | 1512 | 918 | 1512 | 918 | 1512 | 918 | 1512 |
| 1513 | Processing/preserving of fruit & vegetables | 1513 | 8492 | 1513 | 8492 | 1513 | 8492 | 1513 |
| 1514 | Vegetable and animal oils and fats | 1514 | 2723 | 1514 | 2723 | 1514 | 2723 | 1514 |
| 1520 | Dairy products | 1520 | 7554 | 1520 | 7554 | 1520 | 7554 | 1520 |
| 153 | Grain mill products; starches; animal feeds | 153 | 5798 | 153 | 5798 | 153 | 5798 | 153 |
| 1531 | Grain mill products | 1531 | .. | 1531 | .. | 1531 | | 1531 |
| 1532 | Starches and starch products | 1532 | .. | 1532 | | 1532 | | 1532 |
| 1533 | Prepared animal feeds | 1533 | 1640 | 1533 | 1640 | 1533 | 1640 | 1533 |
| 154 | Other food products | 154 | 40148 | 154 | 40148 | 154 | 40148 | 154 |
| 1541 | Bakery products | 1541 | 29844 | 1541 | 29844 | 1541 | 29844 | 1541 |
| 1542 | Sugar | 1542 | 1825 | 1542 | 1825 | 1542 | 1825 | 1542 |
| 1543 | Cocoa, chocolate and sugar confectionery | 1543 | 4786 | 1543 | 4786 | 1543 | 4786 | 1543 |
| 1544 | Macaroni, noodles & similar products | 1544 | 564 | 1544 | Propagated | | 564 | 1544 |
| 1549 | Other food products n.e.c. | 1549 | 3129 | 1549 | 31 | | 3129 | 1549 |
| 155 | Beverages | 155 | 17144 | 155 | 17144 | 155 | 17144 | 155 |
| 1551 | Distilling, rectifying & blending of spirits | 1551 | 2587 | 1551 | 2587 | 1551 | 2587 | 1551 |
| 1552 | Wines | 1552 | 5520 | 1552 | 5520 | 1552 | 5520 | 1552 |
| 1553 | Malt liquors and malt | 1553 | 3158 | 1553 | 3158 | 1553 | 3158 | 1553 |
| 1554 | Soft drinks; mineral waters | 1554 | 5879 | 1554 | 5879 | 1554 | 5879 | 1554 |
| 16 | Tobacco products | 16 | 7336 | 16 | 7336 | 16 | 7336 | 16 |
| 1600 | Tobacco products | 1600 | 7336 | 1600 | 7336 | 1600 | 7336 | 1600 |
| 17 | Textiles | 17 | | 17 | | 17 | | 17 |
| 171 | Spinning, weaving and finishing of textiles | 171 | 13283 | 171 | 13283 | 171 | 13283 | 171 |
| 1711 | Textile fibre preparation; textile weaving | 1711 | 12970 | 1711 | 12970 | 1711 | 12970 | 1711 |
| 1712 | Finishing of textiles | 1712 | 313 | 1712 | 313 | 1712 | 313 | 1712 |
| 172 | Other textiles | 172 | 5894 | 172 | 5894 | 172 | 5894 | 172 |
| 1721 | Made-up textile articles, except apparel | 1721 | 3037 | 1721 | 3037 | 1721 | 3037 | 1721 |
| 1722 | Carpets and rugs | 1722 | 995 | 1722 | 995 | 1722 | | 1722 |

| Year | Table | ISIC | Error |
|---|---|---|---|
| 2005 | 13 | 221 | Incorrect 3-digit sum. Expected: 287 |
| 2005 | 19 | 153 | Incorrect 3-digit sum. Expected: 16 |
| 2005 | 19 | 172 | Incorrect 3-digit sum. Expected: 25 |
| 2005 | 19 | 221 | Incorrect 3-digit sum. Expected: 70 |
| 2005 | 19 | 241 | Incorrect 3-digit sum. Expected: 78 |
| 2005 | 21 | 153 | Incorrect 3-digit sum. Expected: 13 |
| 2005 | 21 | 154 | Incorrect 3-digit sum. Expected: 107 |
| 2005 | 21 | 172 | Incorrect 3-digit sum. Expected: 12 |
| 2005 | 21 | 2310 | Non-numerical value entered |
| 2005 | 21 | 2330 | Non-numerical value entered |
| 2005 | 21 | 3591 | Non-numerical value entered |

*Figure 7: Metadata in the Collection Phase: the Excel questionnaire is read in into the system and the statistician is provided with a range of tools for validation, correction, editing or entering of metadata, etc.*

| | Country | Rev | Note Class | UTable | From | To | ISIC | Stage | Scope | Text |
|---|---------|-----|-----------|--------|------|-----|------|-------|-------|------|
| ▶ | 100 | 3 | ASUP | | | | | | XX-- | [MDHC03505] National Statistical Institute of the Republic of Bulgaria |
| | 100 | 3 | BSOU | | | | | | XX-- | [MDHC01023] Census; survey. |
| | 100 | 3 | CDEV | | | | | | XX-- | [MDHC01728] Data have been converted from NACE (Revision 2.2) |
| | 100 | 3 | DREF | | | | | | XX-- | [MDHC00904] Calendar year. |
| | 100 | 3 | ESCO | | | | | | XX-- | [MDHC06157] All enterprises. |
| | 100 | 3 | FENU1 | | | | | | XX-- | [MDHC05468] Mail questionnaires. |
| | 100 | 3 | FENU2 | | | | | | XX-- | [MDHC05807] Complete enumeration. |
| | 100 | 3 | GADJ | | | | | | XX-- | [MDHC05510] Yes. |
| | 100 | 3 | JCON | | | | | | -X-- | [MDHC05511] _Output_ refers to gross output. |
| | 100 | 3 | JCON | 13 | | | | | X--- | [MDHC04600] Output refers to gross output. |
| | 100 | 3 | JCON | 19 | | | | | X--- | [MDHC05194] Value added is gross value added in basic prices. |
| | 100 | 3 | JCON | 21 | | | | | X--- | [MDHC02864] Gross fixed capital formation refers to the net increase |
| | 100 | 3 | KPUB | | | | | | XX-- | [MDHC04818] Statistical Yearbook of the Republic of Bulgaria, publis |

*Figure 8: Metadata in the Collection Phase: a grid representing all metadata entries to the selected context (country, ISIC Rev., year, etc.). A double click on a row will open an edit dialogue.*



*Figure 9: Metadata in the Collection Phase: a dialogue for editing a particular metadata entry.*

**Transformation/Processing**

34. The data collected by UNIDO from the NSOs and further transformed according to the quality requirements in the transformation phase constitutes the major source of data for several recurrent publications produced by PCF/RST/STA. The metadata collected from the NSOs together with the data undergoes the same transformation process as the data and is complemented by metadata generated by the transformation process. All resulting metadata, including the necessary structural metadata, are used in the dissemination process:

35. The data collected from the primary sources are further transformed to a ready-to use data set. The data transformation is done in five stages, which not only constitute an operational framework for UNIDO statisticians, but also provides additional description of statistics (generated metadata which is attributed to each data item) to users. For details about these stages see UNIDO (1996), pp 6-8, only a brief summary is in order:

     i. Manual detection and if possible correction of obvious reporting errors. The data are kept in original form (Stage 1 data). These data are used for pre-filling the following edition of the questionnaire for the particular country;

     ii. Inconsistent data are corrected using supplementary information from national publications (Stage 2 data). Stage 1 and Stage 2 data are considered as official;

     iii. Data are adjusted to eliminate the departures from the level of ISIC aggregation using national and international sources or supplementary data (Stage 3);

     iv. Missing data are estimated by UNIDO statisticians applying related proportion or interpolation whenever applicable (Stage 4) and

     v. Provisional estimates are made for the latest year (Stage 5).

36. At the same time Source and *Method* metadata are maintained for each data item. If appropriate, re-description of the provided metadata from viewpoint of international comparability is performed.

37. During the processing period a range of descriptive metadata also requires updating such as country names, national currencies and country groups. For example in 1990's after the fall of USSR and break-up of Yugoslavia, a number of new sovereign states emerged in Euro-Asia region. On the other side 12 EU member countries adopted common currency Euro replacing the previous national currencies. More recent changes were related to the democratic republic of Timor Leste and Republic of Montenegro, also recently two more countries joined EU (Bulgaria and Romania) and two countries (Malta and Cyprus) adopted the Euro as national currency.

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| **Metadata used** | | | | | |
| Definitional metadata | These are metadata serving as identifiers and descriptors of the data | See *Table 2* | See *Table 2* | | An important step in the processing phase is to update these metadata as necessary before starting the dissemination |
| Descriptive and Methodological metadata | Describe the deviations from the international standards and the methodology used to acquire the data | Supplier of information, Major deviations from ISIC, reference period, Concepts and definitions of variables, etc. | Created by NSO, but can be re-described/rearranged by UNIDO in order to comply with the international standards | Increase the relevancy of the data and add to the interpretability of the data | |
| | | | | | |
| **Metadata produced (output)** | | | | | |
| Operational metadata | are generated throughout the process of data transformation | Stages, sources, methods | Generated during the processing phase and attributed to each particular data item | | |
| Implicit metadata | these are metadata arising from special application of other metadata | Typical example are the ISIC combinations | Created by NSO, but can be generated also during the processing phase | | |
| | | | | | |

*Table 4: Metadata used (input) and produced (output) in the Data Transformation phase*

**Dissemination**

38. The data collected by UNIDO from the NSOs and further transformed according to the quality requirements in the transformation phase constitutes the major source of data for several recurrent publications produced by PCF/RST/STA. The metadata collected from the NSOs together with the data undergoes the same transformation process as the data and is complemented by metadata generated by the transformation process. All resulting metadata, including the necessary structural metadata, are used in the dissemination process:
    o  To define the dissemination products – for this purpose are used the structural metadata like country names and codes, currency names and codes, classifications, etc.;
    o  To guide the dissemination process – for example the selection of data to be published in the different products depends on the degree of confidence they deserve as identified by the stage (metadata generated in the transformation process);

o   To provide users with the information they may need to interpret the disseminated data.

39. The **International Yearbook of Industrial Statistics** is the main UNIDO statistical product, which has been the most important medium of data dissemination for many years. The latest yearbook released in 2008 covered the data for the period from 1995 to latest year. The country data was updated for 74 countries and is compiled from the Stage 1 and Stage 2 (as described elsewhere in this document).

40. Another medium of UNIDO data dissemination are **CD products**, which might include data from all stages described earlier. The demand of CD products is increasing every year from national and international institutions, academia and researches. For information on purchasing procedures and licensing the readers should refer to www.unido.org/statistics. The latest release of the CD products in 2008 covered the following statistics as shown in *Table 5*.

| CD Product | Classification Level | Number of Countries | Period covered |
|---|---|---|---|
| INDSTAT 4 2008 | 4-digits of ISIC rev-2 | 116 | 1977-2005 |
| | 4-digits of ISIC rev-3 | 117 | 1985-2005 |
| IDSB 2008 | 4-digits of ISIC rev-2 | 81 | 1981-2006 |
| | 4-digits of ISIC rev-3 | 80 | 1990-2006 |
| INDSTAT 2 2008 | 2 digits of ISIC rev-3 | Not yet released for publication | |

*Table 5: UNIDO CD products and their data coverage – edition 2008*

41. Another form of data dissemination is providing statistics by selected variables from the different UNIDO databases for each member state which are posted in UNIDO web-site http://www.unido.org/statistics under the item Country Statistics. Country data in the web site are presented for several years together with the figures for the world and region for comparison over time as well as in relation to the region level.

42. Apart from the recurrent publications listed above, industrial statistics data can be disseminated on **ad-hock queries** mainly for internal but in some cases also for external users.

43. In the following *Figure 10*, *Figure 11*, *Figure 12* and *Figure 13* are shown examples of metadata shown in the different dissemination products.

Figure 10 screenshot annotations:

**M indicates metadata related to a country [object]**

**2001 indicates metadata related to a year [time]**

**[1532A] indicates ISIC combination - implicit metadata**

**M indicates metadata related to a variable**

*Figure 10: Metadata in the Dissemination Phase: different types of metadata visible in the data viewer of the CD product INDSTAT4*

Figure 11 screenshot annotations:

**The link indicates an ISIC combination**

Please note :
- Some ISIC categories shown may include values of other (not shown) categories. Clicking the highlighted/underlined figure will display those categories.
- An asterisk (*) next to the year denotes estimated figures.

*Figure 11: Metadata in the Dissemination Phase: different types of metadata visible in the Web Country Statistics*

**Australia**

**Supplier of information:**
Australian Bureau of Statistics, Canberra.
Industrial statistics for the OECD countries are compiled by the OECD secretariat, which supplies them to UNIDO.

**Basic source of data:**
Annual manufacturing survey.

**Major deviations from ISIC (Revision 3):**
Data have been compiled on the basis of the Australian and New Zealand Standard Industrial Classification (ANZSIC). This classification is not fully compatible to the ISIC (Rev. 3), especially at a high level of disaggregation.

**Reference period:**
Fiscal year ending 30 June of the year indicated.

**Scope:**
All establishments with paid employees.

**Method of data collection:**
Questionnaires are distributed by mail.

**Type of enumeration:**
Not reported.

**Adjusted for non-response:**
Yes.

**Concepts and definitions of variables:**
Number of persons engaged is not converted to full-time equivalent.
Wages and salaries includes salaries and fees of directors and drawings from profits by working proprietors and partners but excludes payments to consultants, contractors and persons paid solely by commission. Payments and benefits in kind are excluded.
Output is measured as the sum of the value of sales of goods or services produced by the unit, receipts for industrial work done or industrial services rendered to others, resales and net change in stocks of goods purchased for resale in the same condition as received, net change in stocks of finished goods and in stocks of work in progress manufactured by the unit, value of electricity sold and value of capital assets manufactured for own use. Advertising incomes are also included. Valuation is made at basic prices.
Value added is output less value of industrial materials and services for processing, value of fuels and purchased electricity, cost of industrial services received, and payments for contract and commission work. Valuation is made at basic prices.

**Related national publications:**
Manufacturing Industry, published by the Australian Bureau of Statistics, Canberra.

*Figure 12: Metadata in the Dissemination Phase: methodological metadata shown in the Yearbook*

Australia

| ISIC | Industry | Note | Number of enterprises (number) 2002 | 2003 | 2004 | 2005 | Note | Number of persons engaged (thousands) 2002 | 2003 | 2004 | 2005 | Note | Wages and salaries (Australian Dollars) 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2930 | Domestic appliances n.e.c. | | ... | ... | ... | ... | | ...n/ | ...n/ | ... | ... | | ... | ... | ... | ... |
| 3000 | Office, accounting and computing machinery | | 858 | 895 | 941 | 867 | | 38.6 | 32.0 | ... | 3.6 | | ... | ... | ... | ... |
| 3110 | Electric motors, generators and transformers | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3120 | Electricity distribution & control apparatus | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3130 | Insulated wire and cable | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3140 | Accumulators, primary cells and batteries | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3150 | Lighting equipment and electric lamps | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3190 | Other electrical equipment n.e.c. | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3210 | Electronic valves, tubes, etc. | | ... | ... | ... | ... | | 21.1p/ | 21.6p/ | ... | ... | | ... | ... | ... | ... |
| 3220 | TV/radio transmitters; line comm. apparatus | | ... | ... | ... | ... | | ...p/ | ...p/ | ... | ... | | ... | ... | ... | ... |
| 3230 | TV and radio receivers and associated goods | | ... | ... | ... | ... | | ...p/ | ...p/ | ... | ... | | ... | ... | ... | ... |
| 331 | Medical, measuring, testing appliances, etc. | | 2824 | 2838 | 2934 | 2905 | | 15.3q/ | 13.8q/ | ... | 15.0 | | ... | ... | ... | ... |
| 3311 | Medical, surgical and orthopaedic equipment | | 1916 | 1885 | 1960 | 1909 | | ... | ... | ... | 8.8 | | ... | ... | ... | ... |
| 3312 | Measuring/testing/navigating appliances,etc. | | 908 | 953 | 974 | 996 | | ... | ... | ... | 6.2 | | ... | ... | ... | ... |
| 3313 | Industrial process control equipment | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3320 | Optical instruments & photographic equipment | | ... | ... | ... | ... | | ...q/ | ...q/ | ... | ... | | ... | ... | ... | ... |
| 3330 | Watches and clocks | | ... | ... | ... | ... | | ...q/ | ...q/ | ... | ... | | ... | ... | ... | ... |
| 3410 | Motor vehicles | | ... | ... | ... | ... | | 72.2r/ | 75.4r/ | ... | ... | | ... | ... | ... | ... |
| 3420 | Automobile bodies, trailers & semi-trailers | | ... | ... | ... | ... | | ...r/ | ...r/ | ... | ... | | ... | ... | ... | ... |
| 3430 | Parts/accessories for automobiles | | ... | ... | ... | ... | | ...r/ | ...r/ | ... | ... | | ... | ... | ... | ... |
| 351 | Building and repairing of ships and boats | | 2434 | 2536 | 2695 | 2825 | | 31.8s/ | 33.6s/ | ... | 15.6 | | ... | ... | ... | ... |
| 3511 | Building and repairing of ships | | 626 | 645 | 640 | 685 | | ... | ... | ... | 7.3 | | ... | ... | ... | ... |
| 3512 | Building/repairing of pleasure/sport. boats | | 1808 | 1891 | 2055 | 2140 | | ... | ... | ... | 8.3 | | ... | ... | ... | ... |
| 3520 | Railway/tramway locomotives & rolling stock | | ... | ... | ... | ... | | ...s/ | ...s/ | ... | ... | | ... | ... | ... | ... |
| 3530 | Aircraft and spacecraft | | ... | ... | ... | ... | | ...s/ | ...s/ | ... | ... | | ... | ... | ... | ... |
| 359 | Transport equipment n.e.c. | | 279 | 282 | 295 | 318 | | ...s/ | ...s/ | ... | 0.9 | | ... | ... | ... | ... |
| 3591 | Motorcycles | | 279 | 282 | 295 | 318 | | ... | ... | ... | 0.9 | | ... | ... | ... | ... |
| 3592 | Bicycles and invalid carriages | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3599 | Other transport equipment n.e.c. | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3610 | Furniture | | ... | ... | ... | ... | | 78.1t/ | 73.7t/ | ... | ... | | ... | ... | ... | ... |
| 369 | Manufacturing n.e.c. | | 8241 | 7959 | 7679 | 7600 | | ...t/ | ...t/ | ... | 21.2 | | ... | ... | ... | ... |
| 3691 | Jewellery and related articles | | 2062 | 1981 | 1856 | 1913 | | ... | ... | ... | 3.9 | | ... | ... | ... | ... |
| 3692 | Musical instruments | | 4777 | 4605 | 4470 | 4338 | | ... | ... | ... | 14.1 | | ... | ... | ... | ... |
| 3693 | Sports goods | | 1402 | 1373 | 1353 | 1349 | | ... | ... | ... | 3.2 | | ... | ... | ... | ... |
| 3694 | Games and toys | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3699 | Other manufacturing n.e.c. | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3710 | Recycling of metal waste and scrap | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| 3720 | Recycling of non-metal waste and scrap | | ... | ... | ... | ... | | ... | ... | ... | ... | | ... | ... | ... | ... |
| D | Total manufacturing | | 129971 | 130116 | 131726 | 131181 | | 1103.8 | 1082.2 | ... | 1069.3 | | ... | ... | ... | ... |

a/ 151 includes 1520, 153, 154 and 155.
b/ 171 includes 172 and 1730.
c/ 1810 includes 1820.
d/ 191 includes 1920.
e/ 2010 includes 202.
f/ 221 includes 222 and 2230.
g/ 2310 includes 2320 and 2330.
h/ 241 includes 242 and 2430.
i/ 251 includes 2520.
j/ 2610 includes 269.
k/ 2710 includes 2720 and 273.
m/ 281 includes 289.
n/ 291 includes 292 and 2930.
p/ 3210 includes 3220 and 3230.
q/ 331 includes 3320 and 3330.
r/ 3410 includes 3420 and 3430.
s/ 351 includes 3520, 3530 and 359.
t/ 3610 includes 369.

*Figure 13: Metadata in the Dissemination Phase: metadata for data elements shown in the Yearbook*

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| **Metadata used** | | | | | |
| Definitional metadata | These are metadata serving as identifiers and descriptors of the data | Country codes and names, Country groups, indicators, classifications: ISIC Rev 2, ISIC Rev 3, currency names and codes | See *Table 2* | | An important step in the processing phase is to update these metadata as necessary before starting the dissemination |
| Implicit metadata | these are metadata arising from special application of other metadata | Typical example are the ISIC combinations | Created by NSO, but can be generated also during the processing phase | | |
| Operational metadata | are generated throughout the process of data transformation | Stages, sources, methods | Generated during the processing phase and attributed to each particular data item | | |
| Systems metadata | drive automated processing | Layout definitions for the yearbook (for each country) as well as country lists, etc., used in the automatic generation of the PDF output; Installation and packaging lists, directories, templates, etc, for creation of the CD product | | | |
| Descriptive and Methodological metadata | Describe the deviations from the international standards and the methodology used to acquire the data | Supplier of information, Major deviations from ISIC, reference period, Concepts and definitions of variables, etc. | Created by NSO, but can be re-described/rearranged by UNIDO in order to comply with the international standards | Increase the relevancy of the data and add to the interpretability of the data | |
| | | | | | |
| **Metadata produced (output)** | | | | | |
| None | | | | | |
| | | | | | |

*Table 6: Metadata used (input) and produced (output) in the Dissemination phase*

# 4. SYSTEMS AND DESIGN ISSUES

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 in-house, taking international standards (ISO/IEC[2] 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.

---

[2] International Organisation for Standardization (ISO) and International Electrotechnical Commission (IEC)

### 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

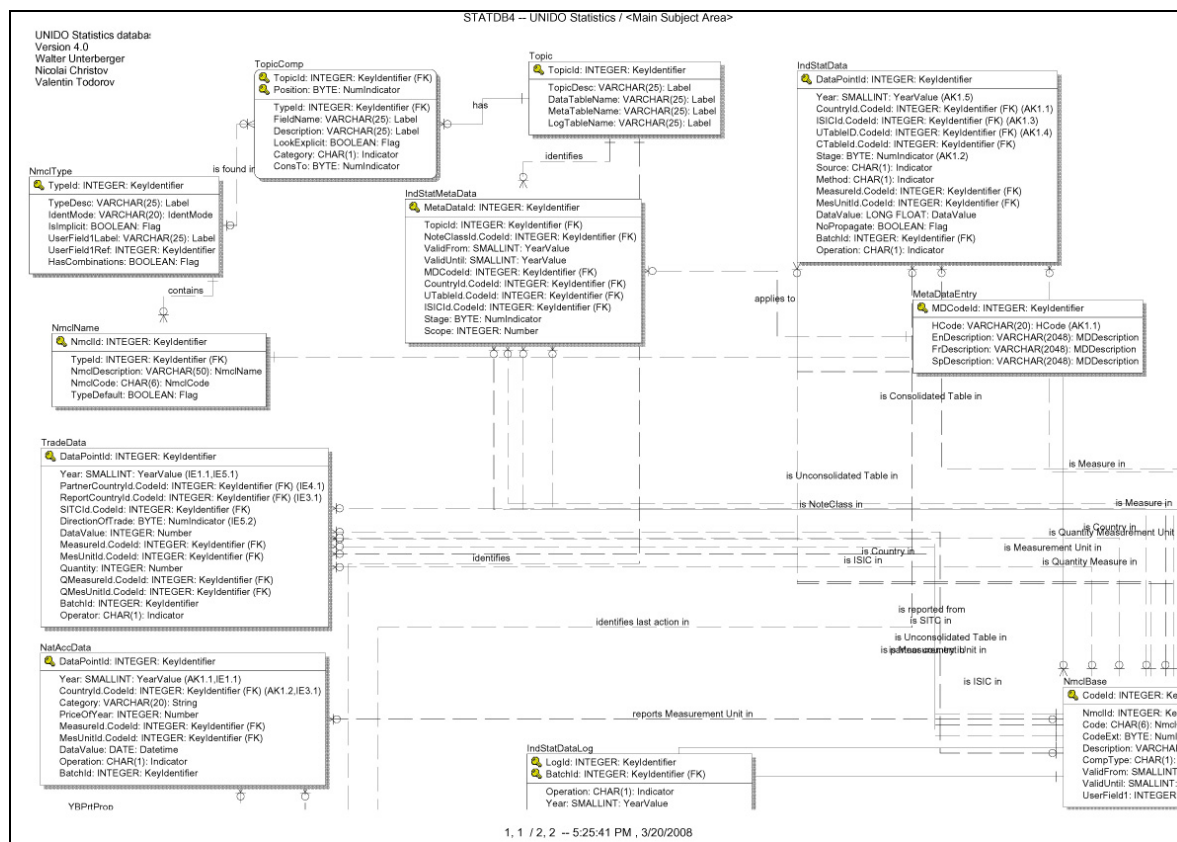50. *Table 1* lists some other tools integrated in the ISDE system.



*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*

# 5. ORGANIZATIONAL AND CULTURAL ISSUES

51. No specialized metadata roles are necessary, since the processing of the metadata is tightly coupled with the processing of that data and the responsibilities are organized by country, i.e. each statistical staff member is responsible for a given number of countries throughout the complete statistical production process (of course the assignment of countries to statisticians is metadata itself and is stored and maintained in the same way as the rest metadata).

52. The introduction of the new Client/Server platform including the new metadata system also did not require any new roles related to the metadata since the same people are maintaining the metadata, but using the modern tools instead of the previously existing clumsy methods.

53. No special training for the staff was necessary since all statisticians participated actively in the specification and the development of the system. As already mentioned, the main part of the system testing was performed by parallel runs on the Client/Server and Mainframe (one very important advantage of the stepwise approach) and the found problems and issues were entered into a simple bug tracking system.

54. A one-week SAS-PC training was given to the staff members in order to facilitate the transition from the mainframe to the Client/Server platform.

# 6. ATTACHMENTS & LINKS

1. Erwin-full.pdf: Data Model of INDSTAT database
2. Malaysia.xls - An example of filled-in Questionnaire in English (2007)
3. Madagascar.xls - An example of filled-in Questionnaire in French (2007)
4. Puerto Rico.xls - An example of filled-in Questionnaire in Spanish (2007)
5. template.xlt – the new questionnaire (metadata) template (2008) in three languages (English, French, Spanish)
6. YBP2-Australia.PDF – An example of metadata used in dissemination – The International Yearbook of Industrial Statistics 2008 for Australia

## REFERENCES:

Fröschl, K.A. and Yamada, T. (2000) The UNIDO Industrial Statistics Information Exchange Architecture (An Integrated Statistical Data and Data Documentation Framework). Working Paper No. 19, Work Session on Statistical Metadata (METIS), Conference of European Statisticians (Washington D.C., 28-30 November 2000).

Fröschl, K.A. and Grossmann, W. (2001) Deciding Statistical Data Quality. In: New Techniques and Technologies for Statistics/Exchange of Technology and Know-how, Pre-Proc. NTTS/ETK 2001, Vol. 1; Hersonissos (Kreta), 2001, pp. 567–575.

Fröschl, K.A. Yamada, T and Kudrna, R. (2002) Industrial Statistics revisited: From Footnotes to Metadata-Information Management, *Austrian Journal of Statistics*, **31** 1, pp. 9-34.

OECD: (2003) Industrial Structure Statistics, Volume 1, Core Data

UNIDO: UNIDO Data Quality (2008), working paper prepared by Shyam Upadhyaya

United Nations (2008) Draft International recommendations for Industrial Statistics, A background document for the 39-th session of the UN Statistical Commission

Yamada, T. (2004) Role of Metadata in Quality Assurance of Multi-country Statistical Data. Paper presented at the Conference on data Quality for International Organizations, Wiesbaden, Germany, 27-28 May, 2004

*** END ***