WP. 28
1 March 2010

ENGLISH ONLY

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Work Session on Statistical Metadata (METIS)
(Geneva, Switzerland, 10-12 March 2010)

ANALYIS OF EXISTING METADATA CASE STUDIES

Submitted by Alice Born (Statistics Canada) and Joza Klep (Statistical Office of the Republic of Slovenia)

BACKGROUND

1. This paper is an update of the analysis of the metadata case studies prepared for the 2009 METIS Workshop in Lisbon. Last year's analysis included metadata case studies from Australia, Canada, Croatia, Czech Republic, New Zealand, Norway, Portugal, South Africa, Sweden and the United Nations Industrial Development Organization. In 2010, metadata case studies from Albania, Austria, Finland, Germany, Netherlands and Slovenia have been added. Latvia will be added when the case study is available. This material is available on the METIS-wiki at http://www1.unece.org/stat/platform/display/metis/.

2. In response to requests from participants at the METIS Work Session in April 2008 and the METIS Workshop in March 2009, the content of these case studies has been analysed and the findings will be presented for discussion at this work session. The analysis has focused on the following sections:

- Section 1 and Section 2.4: strategies for implementation
- Section 4: system and design issues
- Section 5: organizational and workplace culture issues
- Section 6: lessons learned.

3. This paper presents a summary of the content from existing metadata case studies. The findings of the analysis will be given during a separate presentation.

I. STRATEGIES FOR IMPLEMENTATION

1 Overall Metadata Strategy

Albania (new)

One of the main priorities in five years strategic plan named Official Statistical Program; (PSZ 2007-2011) was implementation of metadata in statistics. The main purpose was to inform users of statistical data about:

- which statistical data are available
- where to find and how to retrieve certain statistical data that they need
- how to interpret statistical data, once they are available

To achieve the main purpose, the statistical office decided to follow up the simple strategy as below:

- Provide business orientation as foundation for sharing data assets across entire organization;
- Collection of information about the data collected in an application or database
- Decide how meta data would be used in the organization
- Decide who would use what meta data and why
- Develop map for managing statistical information from different sources
- Improve the quality of statistics produced by INSTAT
- Offer recognition of value of data and its component
- Develop map for managing statistical information from different sources
- Highlight importance of central data administration
- Measurement of value and quality of information

Albanian Statistical Office is in partnership with SCB (represented by International Consulting Office), since 2001. This project is financed from Swedish International Development Cooperation Agency (SIDA). The main purpose of that cooperation is increasing capacities and development of a sustainable statistical system in Albania. Metadata component was put in place in the first phase of project. Furthermore, in second phase of cooperation project, 2006-2009, implementation of metadata system in statistical office was one of the main challenges. The work started with the initiation of development and implementation of a central metadata repository.

Australia

The ABS Metadata Strategy has evolved over the past three decades. It was formalised in an 18 month process, involving stakeholder consultation across the ABS, which culminated at the end of 2003 with the Strategy for End-to-End Management of ABS Metadata being reviewed, and broadly endorsed, by the ABS Executive.

The ABS has now embarked on the path towards another milestone in terms of a 2020 Vision. This is expected to be drawn together during 2009 and then provide a platform for strategic planning over coming years. Ultimately, any ABS metadata strategy exists to support the ABS mission and objectives as set out in the organisation's corporate plan.

In particular, the availability of appropriate metadata and the application of sound metadata management practices are critical to supporting informed use of statistics and the quality of the statistical services the ABS delivers to the nation.

The twelve principles defined as a cornerstone of the 2003 strategy continue to be applied within the ABS:

1. Manage metadata with a life-cycle focus
2. All data is well supported by accessible metadata that is of appropriate quality
3. Ensure that metadata is readily available and useable in the context of client's information need (whether client is internal or external)
4. Single, authoritative source ('registration authority') for each metadata element
5. Registration process (workflow) associated with each metadata element, so that there is a clear identification of ownership, approval status, date of operation etc.
6. Describe metadata flow with the statistical and business processes (alongside the data flow and business logic).
7. Reuse metadata where possible for statistical integration as well as efficiency reasons (no new metadata elements are created until the designer/architect has determined that no appropriate element exists and this fact has been agreed by the relevant 'standards area')
8. Capture at source and enter only once, where possible
9. Capture derivable metadata automatically, where possible
10. Cost/benefit mechanism to ensure that the cost to producers of metadata is justified by the benefit to users of metadata
11. Variations from standards are tightly managed/approved, documented and visible
12. Make metadata active to the greatest extent possible

The strategy proposed the twelve principles be applied when planning and authorising all

ABS projects that provide, and/or make use of, metadata management capabilities, even those where metadata management is a secondary rather than primary objective or requirement.

Other key points in the 2003 strategy include:

- There is an agreed conceptual metadata model which is linked to processes that are part of the statistical processing cycle and this linkage is used to determine what metadata should be collected.
- The ABS metadata model takes account of and uses international standards where possible.
- The physical implementation of the metadata model is the Corporate Metadata Repository (CMR) which is used by all ABS projects. It consists of a number of shared physical databases.
- All metadata entities are managed by a 'registration authority'
- Roles and responsibilities are identified
- Data Management and Classifications Branch (DMCB) is responsible for coordination, definition and maintenance of metadata policies, procedure, systems and provides advice and consultancy to developers related to metadata matters.
- DMCB is the 'registration authority' for the CMR and ensures that other organisational units with this role for particular metadata entities understand that role, are trained and have relevant tools.
- Metadata management is part of every project and should be considered alongside resource allocations and accountabilities in the same way as business processes and data flows are considered.
- Governance of metadata management developments and the oversight of outcomes realisation is vested in line management, existing project and program boards with ABS Executive taking an ultimate corporate view.

Austria (new)

At Statistics Austria, a written and formally adopted corporate metadata strategy does not exist, but the IMS concept could serve as a fundamental cornerstone of such a strategy.

The Statistical Council has explicitly drawn attention to the importance of delivering comprehensive metadata and of increasing the statistical system's coherence, and has demanded the development of a metadata repository. In this context it has also underlined the central role which the IT department should fulfil in implementing the requirements for uniform information delivery, increased quality, enhanced timeliness, easier data access and provision of more comprehensive metadata.

In 2006, an IT project was started (IMS - Integrated Metadata System), the goal of which was to conceive an "integrated metadata repository" based on best practises and international recommendations and to prepare an overall plan for implementing such an information system.

In order to make quick progress in the project and with regard to the limited budgetary and personnel resources, Statistics Austria's top management decided to reduce the scope of the conceptual tasks in the IMS project. The goals and consequently the basic focus of the project were thus stipulated as follows:

"The goal of the system to be developed is to deliver to Statistics Austria's customers (various external users, national and international organisations) that functionality which they require in order to satisfy their needs with regard to statistical information (e.g., to understand statistical results and to have the means to judge their quality). One can start from the assumption that the functional range implemented internationally as "best practise" in various statistical offices will cover the customers' requirements.
Not only external users should profit from the metainformation system. Internal users of statistics also require the metainformation relevant to statistical products and processes (e.g., in order to be able to efficiently reuse statistics produced by the Office or to process

them further for specific projects). It can be assumed that a metadata repository will also generate internal benefits with respect to efficiency and quality of the statistical production process."

With the above as the fundamental goals, the main focus of the IMS project was placed on “passive metadata”, which are required both by external and internal statistics users, in particular for the functions "finding" and "interpreting" statistical data. With an eye to this, the metadata repository was conceived as a comprehensive documentation system for statistical data and production processes.

The concept provides for collecting the metadata which are hitherto scattered over various production systems and (working) documents, storing them in structured form according to a general model of statistics production, and integrating them by allowing links to be created between the individual elements of documentation and the data they describe.

Statistics Austria developed a number of cross-domain applications and stand-alone metadata systems (ISIS, Classification Database, Publication Database, etc.) and which must be connected to the overall system. However, new subsystems must also be developed. These do not replace any existing IT systems, but are responsible for central and structured entry and consolidation of metadata which at present are scattered over various sources. The proposed sub-systems are:
1. Definitions and concepts
2. Statistical projects
3. Types of statistical units and their characteristics (variables)
4. Value domains
The latter two subsystems are based on ISO 11179, although during the modelling process some areas were simplified and others enhanced as compared to this standard.
The subsystems which are to be developed will communicate with each other (and also with the Integrated Metadata Repository IMR, particularly with its component "Registry" - see below) via web services. In this way the mutual interdependencies between the individual subsystems can be minimized, with an eye to the concept of encapsulation.

The Integrated Metadata Repository (IMR) occupies a central position between all these systems. It consists of two parts, the "Registry" and the "Catalogue". The latter contains all those metadata which should be accessible for external users over the Web. These will normally consist of a subset of the metadata administered in the IMS subsystems, as the latter will also contain information which is only of interest to the subject matter persons responsible for statistics production. The Catalogue would also allow a comprehensive full text search over all subsystems, i.e. with a single search request the user should be able to locate not only the documents and Web pages stored in the Publication Database, but also data and metadata in ISIS, the Classification Database and the future IMS subsystems.
The second component of the IMR, the "Registry", is responsible for a Statistics Austria-wide unique registration of all the information objects contained in the individual subsystems and in the connected legacy applications. In addition, it administers links between information objects, which will be of various types (e.g., "contains data from" between a table or an ISIS data cube and one or more surveys). These two core functions are prerequisites for allowing the users to navigate from one data or metadata object to another along predefined paths. E.g., starting from a list of types of statistical units such as enterprise, household, birth, etc., one might quickly locate the characteristics which were collected or created during statistics production, "surf" from there to the corresponding value domains or to the surveys / survey versions, to definitions etc.

Further tasks of the Registry are to provide central services required for more than one subsystem (such as administration of users and access rights, status of the registered information objects, ownership of administered items) and the publication of metadata to the Catalogue.

A dedicated working group (after dicussing the IMS project with top management,

regarding the cost factor and stronger involvement of subject matter statisticians) discussed and specified the contents and functionality of the "Concepts and Definitions" subsystem of the IMS. This system will allow the centralized collection and administration of various definitions relevant to the production of statistics. By integrating the database with the Web content management system, these definitions should be presented to external users on a new metadata portal on Statistics Austria's homepage.

To prevent yet another isolated solution being developed, special emphasis has to be laid on the possibilities of integration and enhancement as they were defined in the architectural design of the IMS. This means that along with the implementation of the "Concepts and Definitions" database the most important components of the Registry and a part of the Catalogue must be realized as well. This should guarantee that - when the system is enlarged by other important modules later on - this can be done without major effort, which means that there will be no obstacles to further development.

Among other results, the working group has elaborated a proposal for a metadata portal and a prototype of the internal metadata management tool (which of course should also be easily extensible). The next step will be to write a detailed requirements analysis document, on the basis of which a precise cost estimate will be made.
If the costs stay within the scope of the available budget, the system will probably be implemented by an external software company.

The working group has also started to debate a second subsystem for the management of metadata which documents statistical projects (as a successor to the standard documentation files). This subsystem should also be able to fulfil EU requirements regarding the documentation of statistics by integrating the Euro SDMX Metadata Structure ESMS.

Canada

One of the primary objectives of the Integrated Metadatabase (IMDB) is to inform users on concepts, methodologies and data accuracy. The IMDB provides the metadata to support the statistical products released by Statistics Canada's Dissemination Division, and relates to the interpretability in the Agency's quality assurance framework.

The responsibility for IMDB and its related *Definitions, data sources and method*s module on Statistics Canada's website was given to Standards Division because it was designed as a means to enforce more coherence in the Agency's data over the long term. The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytical framework, and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys.

This long-term objective of enforcing standardization called for a metadatabase that would store information broken down into building blocks prescribed by a metadata model, where building blocks would be shareable among and linked to all surveys and statistical programs and for any reference period. Where the content of these building blocks evolves over time, storing and linking appropriately versions of them is required. As well, if the metadatabase is to play its role relative to coherence, it has to possess a rigorous registration process to track the validation steps of the information being stored. These are the functions performed by the IMDB.

Croatia

There were three principal reasons to implement metadata in the Croatian Central Bureau for Statistics (CBS):

- to standardize definitions across all statistical activities
- to move the production of statistics closer to the subject-matter experts in order to speed up the statistical survey life cycle
- to present statistics on internet along with its context in order to make statistics understandable and available to users of all types, i.e. to extend the use of statistics

beyond the usual statistical publications

The strategy document was prepared in CBS already in 2001(Zdenko Milonja: Information System Development Strategy, CBS, July 2001). In 2002, a framework agreement was signed between the Division for Western Balkans at the Swedish International Development Cooperation Agency (Sida) and Statistics Sweden's International Consulting Office (ICO). In this framework the Swedish Statistical Office (SCB) provided support for the creation of the public macro database and a central metadata repository in CBS. In its final phases (2006 - 2007) the project was extended to support the development of the Integrated Statistical Information System (ISIS).

Czech Republic

A redesign of the Statistical Information System (SIS) was launched in the CZSO in 2004. The project is centrally managed and monitored by the top management of the CZSO. The first important step was design of a new SIS architecture. The architecture copes with increasing users´ requirements, both on national and international level. It allows effective acquiring and completing of statistical data and metadata.

Major goals for Redesign SIS:

- reducing response burden and boosting respondent motivation;
- optimising production of statistical information in the CZSO;
- designing a conceptual model of SIS and Statistical Metainformation System (SMS);
- defining a unified architecture of statistical tasks. Statistical task - is a set of statistical activities needed to fulfil a user's request for statistical information.
- improving quality of statistical information;
- increasing users' comfort.

The SIS model encompasses a statistical production process (SPP) in all its phases, starting from assessment of users' requirements up to the dissemination of statistical information (SI). The basis for the model is a life cycle of statistical tasks (LCST).

Core principles for Redesign SIS are as follows:

- systematic assessment and evaluation of statistical data requirements,
- increasing share of administrative data,
- increasing use of data modelling,
- implementation of SMS,
- implementation of statistical data warehouse,
- freeze of statistical surveys for 2-3 years,
- avoiding redundancy in statistical surveying.

Finland (new)

Statistics Finland has not laid down a specific metadata strategy, but a policy definition on the development of a centralized statistical metadata system is included in the agency's ICT-strategy: Statistics Finland intends to develop and implement an xml-based common statistical metadata system in order to rationalize and support the harmonization of statistical business processes. The system will be based on Statistics Finland's CoSSI metadata model.

The ICT-strategy and its future goals lay emphasis on creating common and integrated metadata systems application tools. In accordance with the main principle metadata will be created and maintained in a metadata system and made available and transformable to whatever use it is needed for in the statistical business process from data collection to data dissemination.

Germany (new)

Metadata management has been an issue in the statistical system in Germany for many years. Maybe typical for a federal system, solutions have been found and implemented in isolated areas but they have not been coordinated through a common strategy. The current situation therefore resembles a "bottom-up" approach rather than a unified "top-down" solution.

The experience at Destatis and in the Verbund, however, shows that there is a strong need for a more coherent approach to handling metadata in the future. Several key projects in the Verbund - like standardisation of production or quality management - depend on standardised structures and concepts to understand the content of the different statistical activities in a coherent and uniform way. A metadata strategy would also help to provide a framework for the different projects that the metadata section at Destatis is currently working on and that have proven to be difficult to manage as independent projects.

Any future metadata strategy would need to be formulated in accordance with at least the most important stakeholders and it would need to be approved by the responsible committees. Therefore, it is not likely to take shape and become formally adopted in the near future. Meanwhile, there are several projects - independently planned and implemented - that involve a centralised metadata management. Taken together, these projects form the work plan for the next 2-3 years. The task is to combine the projects in a way that at least the outline of a common metadata strategy starts to emerge.

Netherlands (new)

Within the framework of the new business architecture (BA) data and metadata should be stored once and reused as much as possible. The aim is to provide for an office-wide service for the storage and the retrieval of data and metadata. This does not mean that all data is stored: only data with a certain quality meant for general use will be stored and described with the help of metadata. This is what is called steady-state data. Other metadata systems should use the central service as their source for steady-state data.

Concerning the quality of metadata, there is an additional strategy. The core metadata for the central service consists of metadata that is used internationally, preferably EU-metadata. All other metadata should be formulated in terms of this core.

New Zealand

Like many National Statistical Offices around the world, Statistics New Zealand faces a number of 'external' and 'internal' challenges in the years ahead. 'External' challenges include: the need to minimise respondent burden, improve timeliness of existing data releases, improve 'time to market' for new data releases, increased use of administrative data, and better access to data (incl. micro-data) by users. While 'internal' challenges include: provide a better work environment for staff, replace an ageing IT platform & application toolset, measure 'value for money' for the New Zealand tax-payer, develop a platform to support future growth.

In response to these growing demands, Statistics New Zealand developed its Business model Transformation Strategy (BmTS).

The Business Model Transformation Strategy (BmTS) is designing a metadata management strategy that ensures metadata:

1. fits into a metadata framework that can adequately describe all of Statistics New Zealand's data, and under the Official Statistics Strategy (OSS) the data of other agencies
2. documents all the stages of the statistical life cycle from conception to archiving and destruction
3. is centrally accessible
4. is automatically populated during the business process, where ever possible
5. is used to drive the business process
6. is easily accessible by all potential users
7. is populated and maintained by data creators
8. is managed centrally

Established principles of metadata management

- metadata is centrally accessible
- metadata structure should be strongly linked to data

- metadata is shared between data sets
- content structure conforms to standards
- metadata is managed from end-to-end in the data life cycle.
- there is a registration process (workflow) associated with each metadata element
- capture metadata at source, automatically (where possible)
- establish a cost/benefit mechanism to ensure that the cost to producers of metadata is justified by the benefit to users of metadata
- metadata is considered active
- metadata is managed at as a high a level as is possible - managing at the lowest level is prohibitive
- metadata is readily available and useable in the context of client's information needs (internal or external)
- tracking the use of some types of metadata (eg. classifications).

Norway

Statistics Norway has in the course of time developed many different metadata systems. This led to the same information being stored several times in several places making the availability of updated and consistent information difficult. In recent years, there has been a strong focus on the need to link existing systems and a requirement that new metadata systems should not be built in isolation. To facilitate this, Statistics Norway developed a metadata strategy, which was approved early in 2005. The strategy focuses on establishing a conceptual framework, clear roles and responsibilities, and a stepwise development involving integration and linkage of systems.

Portugal

The National Statistical System (NSS) consists of:

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

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

- "To guarantee the coordination of the National Statistical System, approving the concepts, definitions, nomenclatures and other technical instruments of statistical coordination" (Law 6/ 89 of 15 April - Diário da Republica 1st series no. 88).
- This duty is carried out by the "Planning, Coordination and Dissemination" Standing Section (PCDSS), which has the power:
  - "...to analyse and approve concepts, definitions, nomenclatures and other technical instruments of statistical coordination of the National Statistical System and to approve regular changes to these documents resulting from work done at the EU or national level"(Structure and Functioning of Statistical Council - SC Deliberation No. 286; 2005).

The job of the Statistics Portugal is to record, refine, coordinate and disseminate official data while taking into account the general guidelines laid down by the Statistical Council. It may also delegate these duties to other public departments, called delegated bodies.
SP has the responsibility to conceive and manage the statistical metadata system of the NSS, having as presumption that the concepts, classifications and other technical instruments of statistical coordination have to be approved by the SC.
The metadata unit coordinates all the work related to the statistical metadata system.

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

The SP centralises the statistical concepts used in its own and the delegated bodies' statistical surveys in a database. These concepts are classified by subject areas and are entered into database with the status of "proposed concept", when they are used for the first time. Groups of new concepts or changes to approved concepts are sent to the SC periodically for analysis and new approval. The SC has working groups by subject area to

analyse them and recommend their approval to the PCDSS. After the approval, their status in the database is changed to "SC-approved concept" and, is of obligatory use whenever applicable.

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

In 2005, the SP submitted to appreciation to the PCDSS a standard format for the methodological document for the NSS's statistical surveys because it was considered to be a coordination instrument. The format was approved and adopted as mandatory in the NSS. By December 2007, 75% of the surveys in the NSS were documented in this format.

Technical approval of surveys

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

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

Metadata dissemination - Statistics website

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

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

Strategy for the metadata system

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

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

- Will support surveys from their design to the dissemination of results;

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

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

- "Implementing an integrated statistical metadata system — In the development of the statistical metadata system organised and coordinated by the SP, it is particularly important to design and implement an integrated classification management system, an integrated concepts management system and an integrated methodological document management system, define a model for a statistical subsystem and create support instruments for their implementation."
- "Promoting the use of the statistical metadata system in the NSS."
- "Improving user access to statistics - ...adapting the metadata system as a tool for accessing available information and making it easy to read and understand..."
- Two courses of action in the "General Guidelines on National Statistical Activity for 2008-2012" are devoted to the metadata system:
- "To align the statistical metadata system with best international practices."
- "To render the statistical metadata system appropriate to the needs of the interchange of metadata within the National Statistical System and the European Statistical System."

Slovenia (new)

General strategy:
Where standards, identifiers, procedures and tools are developed, their use is mandatory within the office.
SORS corporate metadata strategy was not published as a single document. However, the authors would like to believe that the main goal of SORS in the field of metadata is to develop an efficient and effective, standardized and integrated system for collecting and editing metadata as an important part of the statistical information system.

South Africa

Statistics South Africa's development of the metadata management system has its origins in the organisation's requirement to develop a data warehouse. The idea of a data warehouse came about because the organisation wanted to improve the quality of the statistics produced. It was believed that the data warehouse would play a major role in positioning the organisation within its vision of becoming the "preferred supplier of quality statistics". To begin our data warehouse initiative, we paid exploratory visits to various statistical organisations that had embarked on data warehouse developments in order to learn from their experiences. These visits taught us that a number of things about the complexities, difficulties and peculiarities of developing a data warehouse. In particular, our visit to the Australian Bureau of Statistics showed us that for a data warehouse to have any chance of succeeding in a statistical organisation, it needs to have a strong foundation of standards and policies that govern the statistical production processes. Standardisation of concepts and their definitions, as well as classifications of the terms of the actual survey process, were all found to be necessary for the production of quality statistics. For it to be successful, a data warehouse also needs to operate in this environment.

A formal process for standardization was developed through consultation with standards experts.

The next step for us was to investigate the strength of our standards and policy foundation. Upon this investigation, a number of gaps were identified. Chief among these was the lack of standard metadata in the organisation. The need for standardisation of metadata

necessitated the development of a metadata management system. However, this had to form a good mix with all the other identified ingredients necessary for the production of quality statistics.

Strategically, our metadata management system forms part of a larger system of applications called the End-to-end Statistical Data Management Facility (ESDMF). As an end-to-end system, the ESDMF will consist of tools and applications to support the whole statistical production process. Within this facility exists a metadata subsystem, which plays a central role as the ESDMF was conceived to be metadata driven. In a statistical organisation, a metadata driven system is inevitable because metadata is used and generated at every stage of the statistical production process.

As a data factory, a statistical organisation needs to organise and package data in ways that make it useful to the end user. Produced data must also meet certain minimum quality standards. To satisfy both these requirements, use of metadata is invoked. In packaging its data and statistical products, a statistical organisation must ensure that they are attached with metadata for ease of analysis and interpretation by their users. Metadata also play a key role in ensuring that the end products of this data factory are of good quality. Such metadata includes descriptions of concepts used in the organisation, classifications of these concepts, methodologies and business rules. These are all necessary metadata to ensure that products are of good quality.

The development of a metadata management system was informed by the following principles:

- Maintenance of trust in official statistics: Descriptions of data collection methods, data processing, and storage needed form part how statistical data are presented to the end user. When presented like this, statistical data and products engender trust to the users.
- Facilitation of correct interpretation of statistical data: Metadata accompanying datasets and other statistical products.
- Quality of statistics: Standard metadata contributes to the improvement of a number of quality dimensions. Standardisation of concepts and their definitions and classifications are essential ingredients of standardized metadata.

Programme Providing Frame for Stats SA Projects

The work of all Stats SA components is mapped out in the organisation's Work Programme. Organisational units must support the following strategic themes to advance the work of the organisation:

- Providing Relevant Statistical Information to meet user Needs
- Enhancing the Quality of Products and Services
- Developing and Promoting Statistical Coordination and Partnerships
- Building Human Capacity

This project is aimed at supporting the strategic theme "Enhancing the Quality of Products and Services". Within the DMID project, the metadata management system, more than any of its components, addresses this strategic theme.

Overall Project Objective

Statistics South Africa's metadata management system therefore forms part of the organisation's broader objective to continuously improve the quality of its products. As the driver of the overall facility, the metadata management system is the first deliverables of the DMID project. The metadata management system is also divided into smaller logical units based on the organisation's classification of its metadata. Survey metadata, consisting of elements for providing the overall description of a statistical survey is the first of these metadata deliverables. The survey metadata component is fashioned along the lines of Statistics Canada's Integrated Metadata Database (IMDB) Metastat.

Following the survey metadata component will be the definitional metadata component. This will incorporate into the metadata management system the standardised organisation-

wide concepts and their definitions and classifications as well as other components that form part of definitional metadata.

Sweden (updated)

Statistics Sweden has been active in the metadata field for a long time and has developed several metadata systems templates that contain a lot of metadata. A metadata model for covering the whole production process was created in the late 80's and early 90's (SCBDOK). In the last two years Statistics Sweden has been going through a major reorganization following the Lotta-project. The new way of working is process oriented production with an overall strategy that focuses on customers, efficiency and standardisation. There is however no separate strategy for the metadata management area. Metadata is handled by several organisational bodies within statistics Sweden. Classifications and documentation is by an area in the process department. Process managers are responsible for metadata related to their processes, the communication department is responsible for metadata related to publication in the statistical databases. The Research and development department is responsible for developing the business architecture and its encompassing metadata. The goal is to have common metadata for the whole organisation and that the golden rule of metadata, that it should be collected when it is created is incorporated in the production system in the future.
Metadata development is to be seen as included in the overall strategy for Statistics Sweden's architecture, where metadata has the role of driving and delivering information between different process steps.

There is currently no overall metadata strategy. There is however an unofficial metadata vision, that is in line with the architecture for a Data warehouse and SOA functionality.

UNIDO

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.

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 redesigning the INDSTAT system is its potential applicability to further operational data management areas in need of refashion.

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.

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).

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).

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.

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.

2.4 Implementation Strategy

Albania (new) The project is implemented with a step-wise approach.

The first phase of project served as a preparatory phase. A business structure survey was selected in this phase as pilot. Introduction of this new challenge was a focal point in small team discussions, initiated from developer. Improvements were done step by step. Another survey was used as pilot, short term statistics for a considerable period. The job continue and in the second phase of cooperation project, INSTAT-SCB, 2006-2009 putting it in top priorities. A deadline, till the end of October 2009, was decided for fulfilling the template for all statistical activities carry out from INSTAT and stored in central server. During this year, it is planned to continue developing Metadata Management Tool (AlbMeta Version 2), from the Information Technology Developers. It is planned to merge the Statistical Official Program Database with the AlbMeta Database, and to improve the User Interface.

Australia This question can be viewed from several perspectives. At least in terms of metadata management, the swinging of a pendulum can be seen to some extent.

Developments in the 1990s tended to be on a "big bang" basis. These were sometimes pejoratively referred to as "Cathedral Projects" for being too grandiose in ambition and design, and for taking much longer and much more money to complete than originally expected. Nevertheless, many of the results of these projects have proved to be of enduring value - so much so that many outputs have lived on long beyond their prime.

The strategy next became "opportunistic" and "incremental". There was a broad "master plan" of what should exist in the longer term, but individual "construction projects" were much more modest in scale.

The 2020 vision process on which the ABS has now embarked may move the balance back toward the centre.

At another level, a consistent learning has been that a well developed and managed implementation strategy (in addition to a development strategy) is essential. New capabilities are being delivered into a complex context of existing processes and infrastructure. Uptake of those new capabilities needs to be managed and promoted appropriately. (The simple "Field of Dreams" approach of "Build it and they will come!" has never yet worked for us.) Often the new capability and/or the implementation and communication strategy for it, needs to be refined based on early uptake experience.

Whether it is managed by the development team or some other team, every major project requires a well planned and actively managed "Outcome Realisation" phase after it has finished delivering its major outputs.

A number of reviews in 2008 highlighted the extent to which the complex jumble of existing processes and infrastructure, which lacks transparency and includes many inconsistencies in its underlying information models, is a major barrier both to major change and to progressive evolution in statistical business processes and in supporting infrastructure.

Different processing steps and environments are often joined together through locally designed and maintained "glue steps" which reformat and/or restructure content (data and metadata) to make it suitable for the next step. (For example, if ten different statistical collections perform Process H after Process G there might be ten different customised "glue steps" between G and H - one for each of the collections - rather than a single logically designed "bridge" between the two processes.) "Glue steps" are typically "hardcoded" and often become "black boxes" after the local programmer who built them moves on. However if, for example, Process G changes then the glue steps also need to change (eg F->G and G->H) and there can be further downstream and upstream ripples as well. This adds very significantly to the levels of cost and risk associated with change, creating inertia. This is a major reason why the SESAME Framework described in BHM, or something similar, is seen as a crucial enabler for sustainable change, including implementation of change, in the future.

Austria (new)

Similar to the BASIS 2000+ concept, a modular implementation approach was a major design principle of the IMS.

In order to minimize the complexity of the complete system, the individual components (subsystems) should be able to work independently, communicating with each other and the central "Registry" by means of a web service and program interface layer. Stepwise realization and gradual commissioning and expansion of the IMS should be facilitated.

Regarding the integration of previously existing legacy systems into the IMS, several options are possible. A very simple form of coupling can be realized by manually registering information objects (e.g., a classification from the Classification Database) in the IMS Registry. A tighter and more sophisticated integration will require some programming effort, so that a legacy system can communicate with other components of the IMS via web services.

Canada

The IMDB has been implemented in a "step-wise" approach beginning with 3 development phases, and followed by identifying opportunities to re-use metadata and to expand the IMDB metadata model to link to other information systems in the Agency.

The first phase (Phase 1) was a set of static web pages describing the data sources and methods for each of approximately 400 active statistical programs and surveys, and a similar number of inactive ones. These were accessible only through hyperlinks from data

tables and publications on the Statistics Canada website, while internal users could in addition browse the full inventory of these documents on the Standards Division Intranet site.

In 2000, Phase 2 of the IMDB project was implemented. Metadata was collected from a variety of pre-existing sources and reformatted, submitted to author divisions for validation and loaded into a new metadatabase. A public version was made available on the Statistics Canada website and its internal mirror site while an internal version is posted on the Standards Division Intranet site. The public version can be accessed through hyperlinks from the Daily (i.e., Statistics Canada's vehicle for data releases), and other Statistics Canada products. Updates are triggered by new releases in the Daily so that up-to-date metadata is made available for every new data release. The content of the Phase 2 web pages is based on the requirements of the Policy on Informing Users of Data Quality and Methodology.

Since November 2000, the major efforts have been directed towards improving the quality of the Phase 2 content and developing Phase 3, which adds definitions of concepts, variables and classifications to the database. The Agency continues to complete Phase 3 of the IMDB, which is the most challenging phase to implement since it is the most complex and least documented.

Croatia

The new system will be introduced in a step-by-step manner.

Five surveys were first selected as pilots. The statisticians in charge of these pilot surveys participated in the development teams and provided metadata to CROMETA. Upon presentations and introductions of the pilots new set of surveys will be discussed. Manuals, workshops and training will be provided. Eventually, a deadline will be set for all statisticians to enter metadata into CROMETA using the maintenance tool. The shift of surveys from 'old' to 'new' environment will depend on periodicity and urgency of particular surveys.

Czech Republic

The SMS implementation is, in fact, a 'big-bang' approach.

Introduction of the SMS into practice implies a change in the process of preparing and designing STs by statistical departments. These activities will rely on work with metadata and hence on using SMS tools. A prerequisite for the use of SMS functions is availability of an updated metadata base. What has to be done further is to bring into being all functions and organizational measures related to metadata administration. Adequate training of all participating actors should precede the SMS implementation.

The main condition for introduction of SMS into the SIS operational running is its functionality in all stages of SPP. Effective and viable interlink of SMS subsystems interpreted in a unified metadata base is a necessary precondition for that. This requirement predefines priorities in design and implementation of SMS subsystems implementation strategy.

In view of the project comprehensiveness and complexity, SMS should be developed step by step. The step-wise approach, however, has a clearly defined framework.

The first stage of the SMS introduction into the practice (2008-2009). In the pilot project, Subsystems CLASS, VAR, ST and QUALITY have been tested on the Annual Labour Costs Survey.

The aim of a pilot project:

- definition of ST,
- design of statistical questionnaires,
- data validation (logical control specification),

- design of samples (response duty specifications),
- aggregate specifications,
- output specifications,
- preparation of timetables,
- specification of quality attributes of a statistical task.

The pilot project pre-requires the following:

- to complete a database of statistical classifications (SMS-CLASS),
- to unify methodologically a content of statistical survey(s) for the pilot project,
- to complete a description of statistical variables relevant to the pilot and to ensure their storage in the database (SMS-VAR),
- to create a database for definition of statistical tasks (SMS-ST),
- to develop and test an SMS application program package,
- to develop and make operational statistical data warehouse,
- to establish and make operational an SMS administration,
- to accomplish training of personnel for all professions needed for the pilot project (methodology, subject-matter departments, SMS administration, project preparation, IT applications).

Building up and loading of an SMS database has been for the CZSO an entirely new task. In the newly established SMS-CLASS database, the links to the existing (old) e-system of statistical classifications should be maintained until a complete transition of statistical tasks into the new SIS is accomplished.

The second stage of the SMS introduction into the practice (from 2010 on) will be focused on development, implementation and gradual introduction into practice of SMS subsystems for monitoring of quality, time series, dissemination, respondents and users of statistical information. The second stage will comprise also the completion of SMS-CLASS, VAR, TASKS and D-FUND, namely in terms of their links to the newly prepared SMS subsystems.

Finland (new)

The implementation strategy is step-wise.

The purpose is that once the new metadata system is ready for implementation shifting to its application will happen in parallel with the general modification projects of every statistical data system.

Germany (new)

Since several IT-systems that run on metadata are already in place and given the complexity of the issue, we have decided in favour of a stepwise implementation strategy.

The Statistikdatenbank and the new KlassService are the systems that will become operational first while a variable database and a tool for managing textual documentation (both part of the Census 2011-project) are next in line. A detailed project management is in place in the Census project. The development of KlassService is managed by the Bavarian State Office for Statistics and Data Processing. The progress of the Statistikdatenbank is dependent on the resources of Destatis' IT-department.

The major design work on the Census 2011-related systems will have to be finished in the first half of 2010. As the census will be conducted in May 2011, maintenance and helping users with the systems could have become a major task by then. After 2011 the attention might turn to generalising the lessons learned and broaden the metadata management with involvement in SteP gaining in importance.

Netherlands (new)

The implementation strategy unfolds along multiple lines.

In the first place, all new development projects should act according to the new BA and should take the DSC as a point of departure for the storage of their steady state data. In the second place, existing datasets should be added to the DSC if there is a need for reuse. In

the third place all data arriving from outside the office (the so-called pre-input data) will be stored in the DSC.

New Zealand

The work of the metadata project during 2007 focused on identifying the high level needs of an integrated environment and the adaptation of a metadata conceptual model to understand the relationships and interactions with metadata. With the bigger picture in place, the intention is to focus on developing solutions for smaller components of the wider environment. This approach allows us to focus delivery in the areas which will provide the most gain, while still progressing along the path to delivering the fully integrated solution. It also allows us to assess the strategy at each stage to determine the most practical approach and to minimise the risk of the delivering over-complicated solutions.

Norway

All our metadata projects are based on a step-wise approach.

Portugal

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

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

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

Slovenia (new)

Despite the fact that a large part of the development tasks in the area of information services for the support of statistical production should be finished by the end of 2008, the introduction of solutions into regular production is a particular challenge, and will continue at least until 2010. The introduction of new applications and tools into general use will increase SORS's need for constant improvement of the level of IT services (providing functioning, user support, infrastructure management, etc.) in order for the second main goal of the strategy to be efficient, and that internal and external users of information services are satisfied.

The following actions plans are elaborated in Priorities of the national statistics in 2009:

- Action plan for implementing the strategy for further development of national statistics in Slovenia in 2009;
- action plan for implementing the strategy for further reduction of administrative burdens in 2009;
- action plan for implementing the strategy of quality in 2009;
- action plan for implementing the strategy of data protection in 2009;
- action plan for implementing the strategy of cooperation with reporting units and reduction of burdens in 2009;
- action plan for implementing the strategy for dissemination and communication with users in 2009;

- action plan for implementing the strategy of information technology update in 2009;
- action plan for implementing the strategy of human resource management in 2009;
- action plan for implementing the strategy of financial resources management in 2009.

South Africa no text provided

Sweden (updated) A stepwise implementation approach is used for all systems.

UNIDO no text provided

II. SYSTEM AND DESIGN ISSUES

4.1 IT Architecture

Albania (new)

- database model for the Central Metadata Repository (AlbMeta) is built on MetaPlus from Statistics Sweden
- AlbMeta contains metadata for all statistical activities
- Implemented in MS SQL server 2005 and MS VB . NET for developing maintenance tool

Australia

- service-oriented Architecture (SOA); Oracle Database (store); CRUD services - EIA approach to building corporate systems

Austria (new)

- DB2 on the mainframe computer; web applications run on Linux servers, server programs developed in Java and deployed with IBM Websphere as application server; PC software written in Java
- Number of legacy systems using a number of programming languages, development environments and IT architectures
- If approved, the integrated metadata system (IMS) will be deployed in DB2 on the mainframe, business logic layer as a Web application in IBM Websphere on Linux and client software written in Java

Canada

- "Metaweb" in the Integrated Metadatabase is a web application used to create and maintain metadata content.
- based on a 3-tier thin-client architecture including a 1. presentation tier: browser client, 2. application tier: a web server in the middle and 3. data tier: RDBMS with a database server on the backend (Oracle 10g)
- U.S. Census Bureau's Corporate Metadata Repository (CMR) and ISO/IEC 11179 chosen for the data model

Croatia

- CROMETA server comprises 5 layers. 1. metadata notion generic model is based on the MetaNet Reference Model from EUROSTAT; 2. metadata central storage - MS SQL server platform; 3. metadata business layer implements the conceptual model from layer 1 and the storage specifics in layer 2 - object-oriented approach; 4. metadata consumer services integrates metadata consumers with the metadata server through XML-based web service; 5. metadata manager provides access to the metadata server for adding, browsing, etc. metadata running on MS Windows environment;
- technical platforms include Sybase Power Designer; database is MS SQL server platform and maintenance is done in VB.NET

Czech Republic The global architecture of SIS (GASIS) is composed of three components:
1. Content component of SIS
There is a significant shift in the content component from the statistical survey approach towards statistical object oriented approach. The content component identifies data sources, links between surveys of different periodicity and different purposes; defines modelling

methods, stratification of samples etc. The following three types of statistical variable are distinguished:

- fundamental variable (used for calibration and/or modelling),
- standard variable (predefined set of the statistically most important variables),
- complementary variable (supporting fundamental and standard variables).

2. Metainformation component (SMS)
Systematic use of metainformation inside and outside the SIS as a tool for internal and external integration. SMS is focused on the SPP. The model used for definition of a statistical variable ensures its standard description from the beginning up to the end of LCST.
3. ICT component
Software and hardware support for SPP. Standardisation of application software used in all stages of SPP. Tools for mathematical models and mathematical and statistical methods. Tools for data approval, release and dissemination. Statistical data warehouse and public database.

- Statistical Metadata System (SMS) is integral part of the IT architecture of SIS, and SMS is a precondition for all datawarehouses;
- Technical infrastructure: UNIX servers and Oracle database; work stations in MS Windows environment;
- SMS application: O/S is MS Windows or Linux;
- metadata viewed with an Internet browser;
- future developments are Oracle Forms and Reports (3-layer architecture), Oracle Application Server; Java server pages; for metadata presentation on the Internet, Internet Java Server Pages technology will be used.

Finland (new)

- repositories for metadata including classifications, concepts, data descriptions and variables but no tools for new and updating metadata
- looking at a centralized metadata system for all phases of the statistical process
- in transition from relational database to an xml-based environment – new ICT strategy will be an xml database as the primary metadata respository
- currently, there is a Sybase relational database and a new XML database (eXist) containing the metadata editor
- the relational database is also used form archiving statistical data; classifications and concepts are generated from the DB to web pages and can be used in SAS and Superstar
- first metadata tool linked to XML DB is the variable editor to be implemented in 2010
- statistical data are stored in Sybase and MS SQL, and in eXist XML DB; programming is done in .net, PowerBuilder and SAS; and tabulating tools are SAS and SuperStar
- considering using MetaAPI interface in eXist in order to ensure metadatabases are used in all statistical processing systems through an interface that is not dependent on a particular metadatabase and underlying data models

- there are separate metadata systems but the plan is to have these systems compatible with other common metadata systems

Germany (new)

- no all-encompassing metadata system exists or will exist in the future due to organizational policies; each metadata system has its own IT architecture

- to allow users to access metadata stored in metadata repositories, a web portal will be developed and will be connected to the portal through we service functions

- Neuchâtel model was integrated into existing and new systems to support the 2011 Census

Netherlands (new)

- Based on Enterprise Service Bus with links to a data service centre, a meta service centre and a classification server; and linked to the phases of statistical business processing

New Zealand

- service-oriented architecture – as an enabler; developing a new system – Business-model Transformation Strategy (BmTS) with 3 deveiliverables: 1. generic E2E business process; 2. information architecture; and 3. technical architecture;
- Advantages are: able to use third party software and transparent exchange of metadata between systems;
- current service layer supports existing metadata components – process management, business rules, respondent management, tools (SAS, ETL, Blaise) and databases (SQL Server)

Norway

- service-oriented architecture – support open standards, platform independent and component based, and web services;
- advantages: collaboration between old and new applications; reuse of systems;
- replace systems behind the scenes with no direct impact on users;
- information architecture includes data collection, metadata and dissemination systems

Portugal

- each subsystem in the integrated metadata system has a similar architecture – a database, two Web applications, view that provides metadata to be reused;
- management is decentralized with central co-ordination

Slovenia (new)

- consolidated development environment using relational databases (Oracle and some MS SQL Server); .NET as the application development environment; SAS for processing and analyses; and tools from other NSOs – Blaise, PC-Axis, Argus, etc. for the phases of the statistical process
- main challenge is to integrate individual technologies into one system
- achieved by using service-oriented architecture (SOA)
- will introduce SDMX technological solutions for XML formatted data and metadata
- introduction of ITIL guidelines
- shared servers and dedication ISIS servers

South Africa

- enterprise architecture approach – software architecture of applications in ESDMF (E2E Statistical Data Management Facility) is dictated by the IT architecture;
- follows these architectural principles: 1. integration with other systems; 2. interoperability using Java; 3. modularity of metadata management system by category of metadata; 4. scalability for growth; 5. flexibility of applications – use of object-oriented programming

Sweden (updated)

- structured and well-documented data warehouses
- 3 base registers – individuals, businesses and real estate
- minimize duplicate data and data transport
- support process-oriented production, efficient data collection and flexible dissemination
- contains a stock of variables and time-stamped data

UNIDO

- structure of the Integrated Statistical Development Environment (ISDE) is a 3-tier architecture built in .NET technology;
- data and metadata are stored in a centralized database and user interacts with the system through the ISDE shell

4.2 Metadata management tools

Albania (new)

- metadata management tool is developed in Visual Studio 2005 and VB .NET

Australia

- current approach is an application "reading" relevant content from a metadata repository rather than writing back new or updated records;
- looking at SDMX

Austria (new)

- due to existence of several metadata systems, there are a number of different metadata management tools including:
    - Standard Documentation – predefined templates, replicated on Web server
    - eQuest – electronic questionnaires developed in VB 6.0
    - Classification Database – software for interactive development of new classifications
    - IMS – management tool with extended plug-ins based on Eclipse RCP is planned

Canada

- Metaweb is a web application to create and maintain metadata content – survey metadata and definitional metadata (object class, property, data element and data element concept); Metaweb is Java JSP and Servlet based;
- value domains (classifications) are populated into the database through a MS Excel VB Loader and an Oracle PL/SQL IMDB loader;
- information discovery is through Statcan Wiki where data is extracted from IMDB Oracle database using a VB .NET application

Croatia

- CROMETA metadata manager is used to add new, edit and delete metadata from central repository

Czech Republic

- for communication between SMS subsystems and other applications, direct communication technology between Oracle data tables of SMS and table of other subsystems;
- XML technology for SMS interfaces;
- user interface to SMS subsystems is Oracle Forms;
- data link to metadata will be established in datawarehouse environment from the VAR, CLASS and TASKS metadata subsystems

Finland (new)

- basic tools for metadata systems will be developed in 2008-2012 including a variable editor, classification editor, table editor, process metadata and metaWeb service
- XML database will be the primary repository for metadata replacing relational databases
- currently, metadata are copied from the relational database to the XML database

Germany (new)

- links between data and metadata exist in the .BASE system with a tools for defining data editing rules
- variables and value domains are stored in a separate repository
- GENESIS cube database contains a metadata repository – before a cube is created variables are defined and reuse of existing metadata is supported, the system contains about 1,300 variables

Netherlands (new)

- data and metadata are stored with the help of Documentum
- classifications and code lists are stored with the help of the Classification Server

New Zealand

- developing components of metadata infrastructure including: 1. search and discovery, access/registration; 2. metadata storage including data definitions, passive metadata

store, classification management (new CARS), question/variable library and business logic; and 3. other storage - frame and reference stores, documentation and reports, and standards and processes.

Norway
- Oracle databases with links between databases;
- web applications obtain metadata from service library for metadata systems

Portugal
- based on : 1. operating system (IIS servers, databases using MS O/S); 2. Network architecture supports both LAN and WAN; 3. applications supporting metadata system are Web applications developed in .NET; 4. Storage is MS SQL Server - easier to integrate other systems

Slovenia (new)
- no special tools used

South Africa
- metadata management application is grouped into three tools: 1. administration module - manage users of the system and to make changes to the system; 2. metadata capturing and editing module - captures survey metadata at the survey instance; 3. query and reporting module - to view captured metadata

Sweden (updated)
- MetaPlus contains an administrative tool
- A Super cross database is used for harmonization and outputs using metadata

UNIDO
- 1. database is Sybase ASE 12.5; 2. database maintenance is SQL Programmer 12.0; 3. application development in .NET framework 2.0

4.3 Standards and formats

Albania (new)
- no standards and formats have decided but plan to implement SDMX format

Australia
- Current: Collection Management System; Dataset registry (ISO/IEC 11179); Classification Management System; Data Element Registry (DER) (ISO/IEC11179); Questionnaire Development Tool; Question, Question Modules, Collection Instruments; Quality Infrastructure (QIS) and Business Activity Monitoring (BAM); Process metadata (looking at SDMX and other business process software)

Austria (new)
- Classification database is based on Neuchâtel model
- IMS proposed subsystems (variables and value domains) are based on ISO/IEC 11179

Canada
- CMR and ISO/IEC 11179 Metadata Registries(2003 release) is used for data model;
- data entry in Metaweb is managed through web forms;
- storage of the metadata is in an Oracle database; XHTML pages are dynamically generated with Perl scripts

Croatia
- CROMETA model contains customized Reference Model TM concepts;
- model is evolving to support the needs of statistical processing;
- CROMETA model can be implemented on any technical platform;
- available as UML;
- SDMX standards are going to implemented

Czech Republic
- SMS subsystems use: XML interface or exports to other SMS or SIS applications;
- Oracle Forms 10g can be started at work stations;
- Internet mirror of CLASS created in Java Server Pages;
- backup in Oracle ARCHIVELOG; SDMX for data interchange in the future

Finland (new)
- Common System of Statistical Information (CoSSI) based on XML, which is a modular data model for describing statistical tables, classifications, concepts, variables, general information on statistical documents, quality descriptions, etc.
- CoSSI is compliant with Dublin Core and CALS

Germany (new)
- .BASE runs on nationally designed XML formats – Survey Definition Format (DatML/SDF) describes the survey and its variables; DatML/EDT defines the editing

rules; and DatML/ASK is metadata to create electronic questionnaires

- GENESIS has its own database model – national standard used in regional offices; generates data and metadata in SDMX format
- KlassService will be based on the Neuchâtel model
- No standard format is used in the Statistikdatenbank but is similar to the Neuchâtel model, and can interact with other systems in a distributed metadata system

Netherlands (new)

- metadata standards are incorporated in the SN metadata model – adoption of ISO/IEC 11179
- Documentum uses its own formats and standards, and exports metadata in XML and CSV file formats
- Metadata is produced during the design phase

New Zealand

- four concept-based models were reviewed - DDI, SDMX, MetaNet Reference Model v2.0 and Neuchâtel Terminology Model;
- blended model was selected - MetaNet as the primary model and SDMX as the secondary layer- and is currently under development

Norway

- Classification system is based on Neuchâtel Terminology Model Part 1 Classification v2.0;
- variable system based on partial implementation of Neuchâtel Terminology Model Part 2 Variables with some use of ISO/IEC 11179;
- considering using DDI v3 for exchange of microdata;
- used definitions of metadata concepts from SDMX MCV

Portugal

- Excel, CSV and PDF

Slovenia (new)

- No text provided

South Africa

- Metadata management application comprises: 1. user interface, 2. database, 3. business logic, 4. application/web server

Sweden (updated)

- variables in MetaPlus is based on ISO/IEC 11179
- classification database KDB uses Neuchâtel terminology for classifications

UNIDO

- considers ISO/IEC 11179 metadata registry standard

4.4 Version control and revisions

Albania (new)

- no approach for version control; plans are to customize the approach in MetaPlus

Australia

- looking at the approach set out in ISO/IEC 11179 Part 6 - registration;
- customised versioning rules depending on the system

Austria (new)

- in the IMS, specify version control for certain objects (i.e., survey versions)
- for *concepts and definitions* in the IMS, no version control will be implemented since they are relatively static
- revisions will be a feature in the metadata repository software

Canada

- IMDB uses open source version of control tool Concurrent Versions System (CVS) and Windows Clients for CVS is used for software source code;

	• separate environments are used for development, acceptance testing and production
Croatia	• each metadata object is versioned with only one version valid at any time as determined by the "status" of the object (e.g., under development, released, authorized, archived, frozen or deleted)
Czech Republic	• object state is SMS is stable in a time interval; and in a certain validity range is used the concept of "object version"; • object version each object version may have the following states - under preparation; for approval; approved; revised and revised for approval.
Finland (new)	• 4-level hierarchy for versioning of classifications: level 1 classification name; level 2 statistical versions of a classification; level 3 time versions of a classification; and level 4 language versions of a classification • concepts are versioned the same as classifications
Germany (new)	• two ways for version control: 1. attach validity periods to metadata objects; and 2. create additional object types for the versions of a metadata object type – one for general information that does not change over time, and another for a list of versions to record changes over time • validity periods are used in the .BASE system for active surveys; inactive surveys are disposed or archived • instantiations (instances?) are used in KlassService following the Neuchâtel model where classification versions are their own object type • instantiations are in RDC metadata system for statistical activities and successive surveys
Netherlands (new)	• Documentum has a version control mechanism for changes over time and for changes in the design a datasets
New Zealand	• part of Logical View of data and metadata stores and within the "reference metadata layer"; • versions of data definitions within the "structural metadata layer" and linked to versions of reference metadata
Norway	• metadata have valid to, valid from and last updated; • program code for web services is checked in and out
Portugal	• versioning is managed by the managers of each subsystem with integrity rules
Slovenia (new)	• different rules implemented
South Africa	• edit function in the Survey Metadata tool allows for the revision of metadata by users with permissions
Sweden (updated)	• metadata in MetaPlus have a valid-to-valid-from date stamp with the capability to add time series and historical information • metadata is logged and can be recreated • reference metadata have a reference time
UNIDO	• no text provided

4.5 Outsourcing v.s. in-house development

Albania (new)	• in-house development
Australia	• in-house development; open source and collaborative options will be used in future developments; • interoperability between repositories (internal and external), standards and vendor software in future considerations
Austria (new)	• Project management is done internally and software development will be outsourced • Classification database was developed in-house
Canada	• in-house development; open source
Croatia	• in-house development; given the complexity of CROMETA concepts, outsourcing is not possible
Czech Republic	• technology design will be outsourced; • content specification developed in-house
Finland (new)	• developed in-house
Germany (new)	• combination of in-house development, Verbund development and outsourcing
Netherlands (new)	• SN tries to minimize tailor-made software development • Documentum is a universal tool for storing and retrieving files – including both data and metadata
New Zealand	• strategy of “enhance, then buy before build”; • data definition is being developed in house; • looking at feasibility of enhancing CARS to incorporate question and variable management; redevelop survey metadata tool
Norway	• mainly in-house development of metadata systems and business processes
Portugal	• developed in-house
Slovenia (new)	• use both in-house and external sources
South Africa	• development of the metadata management system (as part of the ESDMF) was outsourced due to lack of specialized resources
Sweden (updated)	• developed in-house • SSD metadata developed within PC-Axis
UNIDO	• developed in-house

4.6 Sharing of software components of tools

Albania (new)	• INSTAT is not in a position to share software or tools at this time
Australia	• systems have been built in a “monolithic” fashion and highly customized for ABS environment; • makes sharing very complex; applications are 10 years old; • DER and QDT are newer and easier to share – user interface to Oracle; • software should be made available free of charge if in current form; • if changes are required, cost recovery basis or collaborative arrangement; exchange details of the data models, etc.;

	• ABS looking for partnerships for new metadata system
Austria (new)	• No text provided
Canada	• IMDB is currently 10 years old and is being upgraded thus making sharing complex; • Statistics Canada is willing to share its documentation on the IMDB model.
Croatia	• no text provided
Czech Republic	• CZSO is willing to share applications subject to discussions between CZSO, EU and contractors
Finland (new)	• Applications developed at Statistics Finland can be shared free of charge with other statistical organizations
Germany (new)	• Major problem with sharing is language since most interfaces are in German only and systems do not support multi-lingual formats, except KlassService supports n-languages and GENESIS supports English and German • Not really possible, however this would need to be reviewed
Netherlands (new)	• SN Classification Server is tailor made and could be used by other organizations; however, is still in the testing phase
New Zealand	• no text provided
Norway	• data models and XML schema are available on request
Portugal	• no text provided
Slovenia (new)	• will to share scripts, guidelines and manuals for classification database, advance calendar release, WeCMS, registry of surveys and survey instances, and planned activities within survey instances
South Africa	• no text provided
Sweden (updated)	• Statistics Sweden can share MetaPlus system and business process model • MetaPlus and its application are available in English
UNIDO	• no text provided

III. ORGANIZATIONAL AND WORKPLACE CULTURE ISSUES

5.1 and 5.2 Overview of roles, responsibilities and metadata management team

Albania (new)	Organizational units and responsibilities: • subject-matter statisticians responsible for content and documentation in SCBDOC template • IT developer responsible for system architecture, web designing and database management • Metadata steering group responsible for co-ordination of work done by the two groups mentioned above
Australia	Staff roles and responsibilities: • For new projects - project planners, project managers, business analysts, IT staff • For established products – subject matter staff. In addition to documenting their metadata initially, senior subject matter staff became responsible for "signing off" that

the documented content was both accurate and sufficient.

- Specialists in metadata management – 9 individuals.

Organisational units and responsibilities: Data Management Section (DMS), three subject matter areas each with a co-ordination section, publishing area and two "standards" sections - Population Statistics Standards (PSS) and Economic Standards and Classifications (ESC).

- Data Management Section (DMS) resides within the Data Management and Classifications Branch (DMCB) of the Methodology and Data Management Division (MDMD) of the ABS.
- DMS consists of around a dozen staff supported by around half a dozen programmers (application developers) from the ABS Technology Services Division
- DMS developed guidelines to assist key personnel in new projects in understanding the practical meaning and intentions of the principles set out in the metadata management strategy and how they might apply in the context of a specific project.
- DMS also provides direct interactive advice to planners, analysts and IT staff.
- DMS looks after policy and strategy related to metadata, the work program related to the Corporate Metadata Repository (CMR) and user support and training related to the CMR
- DMS ensures the necessary "repository infrastructure" is provided, and that the infrastructure remains "fit for purpose" in a changing organizational and technical environment
- As part of supporting the National Statistical Service, DMS also increasingly engages with other agencies in regard to data and metadata management issues. In some cases this takes the form of providing guidance and advice, in other cases the focus is more on bilateral sharing of perspectives, experiences and plans.
- DMS is not responsible for the quality of the content held within each repository.
- DMS also looks after the work program, user support and training for the output data warehouse and other aspects of data management policy and practice within the ABS
- Statistical subject matter areas are required to make appropriate use of the available facilities, adhere to the policies and follow the relevant guidelines. In particular, these areas remain responsible for the extent, accuracy and other aspects of the fitness for purpose of the metadata content related to their particular collection, including classifications, data elements etc which are specific to their collection.
- Subject matter areas also became responsible for ongoing custodianship of that metadata, including ensuring it remains up to date and answering any enquiries its definition might generate from others.
- Each of the three subject matter Groups within the ABS includes a "co-ordination section" that assists with requirements gathering and prioritisation for new metadata facilities and for improvements to existing ones, targeting and coordinating "user acceptance testing" of, and feedback on, new/changed facilities , coordinating definition of implementation programs for new processes and systems, monitoring progress of implementation and escalating the most common and most serious implementation issues to ensure they are addressed, aiding communication between DMS and end users (translating terminology and impacts between the two) and meshing the CMR work program with other work programs relevant to that Group
- The two other sections within DMCB are the standards areas for economic and population statistics, looking after the development, definition and promotion of key content related statistical frameworks, concepts and classifications.
- Population Statistics Standards (PSS) and Economic Standards and Classifications (ESC), have leadership roles in regard to metadata content within the ABS. They develop and support "standard" classifications and data elements with are cross domain in nature (e.g., industry, occupation, language). Many of these are recognised standards for Australia as a whole, not just the ABS. The standards areas also provide guidelines and advice to help subject matter areas ensure their "collection specific" metadata is well defined and curated.
- DMS also works closely with the publishing area to facilitate appropriate content from

the CMR flowing through appropriately into publications or directly onto ABS web pages.

Other relevant groups and responsibilities: Metadata Strategy Group (MSG), Metadata Community of Practice (CoP)

- At a higher level the Metadata Strategy Group (MSG) comprising "Branch Heads" drawn from across the ABS exists to elaborate upon, drive forward and "champion" the metadata strategy. This group has direct access to top level management with the ABS and has regularly brought critical issues and proposals before them for input and funding approval.
- The Metadata CoP will focus on co-ordination of work programs and activities, sharing of leading practice and information exchange

Austria (new)

- No unit is responsible for a centralized metadata system, and these responsibilities are distributed among the various organizational units
- There is a *Registers, Classifications and Methods* department (70 persons) with a GIS unit that is responsible for spatial metadata due to the EU directive INSPIRE
- Total Quality Management Board is responsible for quality of the statistics and conducts quality projects
- IT department have a leading function in conceptual planning and realization of metadata systems

Canada

Staff roles and responsibilities: officers, analysts, database administrators

- The IMDB architectural design and systems development is provided by three officers of the Systems Development Division, which is also part of IMF.
- The chief of the CMS directs the work of three analysts responsible for the development of the content of the base, four officers responsible for the maintenance and update of the content, and one database administrator.

Organisational units and responsibilities: Corporate Metadata Section (CMS) and standards sections in the Standards Division in the Informatics and Methodology Field

- The Integrated Metadatabase (IMDB) is Statistics Canada's corporate metadatabase initially put in place to allow proper interpretability of the data released to public.
- The development and maintenance of the IMDB is the responsibility of Standards Division
- The Corporate Metadata Section (CMS), is dedicated to the IMDB activities
- Three other sections of the division look after the development and maintenance of the Agency's standards (i.e., statistical units, concepts, variables) and classifications.

Croatia

Staff roles and responsibilities: statisticians, application developers / information system developers / programmers

- statisticians are responsible for respective statistical surveys. They manage statistical surveys according the periodicity and they are responsible for the timelines and quality of statistics. Software is developed upon the written requirement issued by the statistician in charge. The requirement document is a template based on the needs of IT experts to process any survey.

Organisational units and responsibilities: 25 subject-matter departments, IT sector

- Data of almost all statistical surveys are handled and processed in the IT sector

Development project: CROMETA has been divided in two parts

- the *Metadata Methodology* sub-project with the task to form and define the conceptual metadata methodology for CBS, based on mapping of all existing metadata models and definition of processes needed to reach a comprehensive metadata solution. The result of this project should be the basis for technical implementation.

- the *Technical Implementation* sub-project with the goal to develop the solution to physical implementation of the conceptual model of CBS' metadata and metadata management tool. The result of this project should be the implementation of the 'empty' metadata container and the user interface.
- Each sub-project had its own development team consisting of 7 - 8 people; between them one project leader, 1 - 2 full time developers, others as supporting members with 30-75% engagements in the respective project. Unfortunately, the project suffered a significant loss of resources, nevertheless the interesting and challenging work. By the year 2007 the metadata repository was practically completed by three developers.

Czech Republic

Staff roles and responsibilities: methodologists, subject-matter statisticians, IT specialists, programmers, specialists on statistical dissemination.

Organisational units and responsibilities: SMS administration

- SMS administration must be established and will be an integral part of the organizational structure of the CZSO.
- SMS administration will fulfill following roles central administration of SMS and all subsystems.

Development project: Project steering committee (PSC), Task Force SMS, top management, project teams

- Top management appoints members of PSC, Task Force SMS and project teams and reviews progress reports from PSC
- PSC makes progress reports on SMS subsystems for top management, continually controls the results of the work (at least once in three months) and takes principal decisions on the way of further work.
- Task Force SMS has the responsibility for administrative coordination/organisation of SMS project teams in cooperation with the heads of project teams for individual SMS subsystems. The work includes the preparation of regular progress reports and creation of conditions for activities of external project workers inside the Office.
- The project teams work on design of individual SMS subsystems and cooperate with external project workers in development, testing and putting the subsystem into the CZSO practice.

Finland (new)

- Guidelines for development of metadata systems are co-ordinated with IT management, Dissemination Services and Classification and Metadata Services
- Classification and Metadata Services maintains classification standards, concepts and the archiving metadata system;
- Statistical departments maintain their own classifications and concepts in centralized metadata systems according to the guidelines of the Classification and Metadata Services unit
- Metadata Co-ordination Group informally meet to work on metadata issues
- CoSSI-model Steering Group is charge of developing the model according to user needs in a manner that will not expose its main structure to risk

Germany (new)

Staff roles and responsibilities:

- Central metadata unit co-ordinates the different units that deal with metadata – metadata used by the output database (GENESIS) is co-ordinated by the dissemination department; metadata that describes microdata files is stored in the RDC metadata system and is jointly maintained by the Destatis and regional RDCs; and production metadata is maintained by subject-matter departments

Netherlands (new)

- apart from general default roles that Documentum provides (e.g., author and owner), there is a metadata administrator responsible for metadata standards, a metadata designer responsible for the design of datasets and supplying correct variable definitions; and a user who is allowed to browse the metadata and statistical data

New Zealand

Staff roles and responsibilities: statistical analysts, IT personnel, management, data managers/ custodians/ archivists, statistical methodologists, architects, respondent

management, survey developers, metadata and interoperability experts, project managers and teams, IT management, product development and publishing, customer/client management.

External user groups: government, public, external statisticians

Other groups and responsibilities: information management group, standards, solutions and capability group

- Teams within our information management group and standards, solutions and capability group are currently involved in developing our metadata processes and infrastructure.
- These groups will also be made responsible for the ongoing training, knowledge management and support of the metadata solution.

Norway

Staff roles and responsibilities: Subject matter statistician, survey manager, metadata manager, senior advisers in standards, IT developers, system architects and web designers

- The system architect for metadata systems has the responsibility to ensure that new development is in line with the metadata strategy and the IT strategy. This is possible through an approval process for system development documents.

Organisational units and responsibilities: Division for Statistical Methods, Standards, the IT Division for Communication Systems and Division for IT infrastructure within the Department of IT and Statistical Methods. Division for IT services in the Department of Data Collection.

- Our core team for development of metadata systems consists of one senior adviser in Division for Statistical Methods and Standards and one system architect for metadata systems and two programmers in the IT Division for Communication Systems
- When necessary we draw on expertise in data capture, web and web-services from the same IT division and on resources from other IT divisions.
- Metadata system maintenance is carried out by two or three people spread across the Division for IT Division for Communication Systems, the Division for IT infrastructure in and the Division for IT services.

  Maintenance of the system contents, i.e. the metadata, is carried out by all statistical divisions with support from the senior adviser in the core metadata team.

Portugal

Staff roles and responsibilities : IT specialists, metadata system manager, metadata subsystem managers, dissemination database manager, technicians, nomenclaturists, terminologists, responsible of implementing the SDMX standard:

- Duties of the metadata system manager are: Coordinating managers of each subsystem in the system; Ensuring that the different subsystems' conceptual models are properly integrated; Defining the general harmonisation rules applicable to all subsystems, in cooperation with the subsystem managers; Planning training courses, subsystem revisions, etc. in cooperation with the subsystem managers.
- Each metadata, concepts, classifications, variables, methodological documents and data collection instruments subsystem has a manager in the Metadata Unit who guarantees the application of standardisation and harmonisation rules in each subsystem. These managers hold talks whenever necessary to articulate coherence and integrity between the different subsystems. They also have discussions with the survey managers and the Dissemination Database manager
- Technicians with a more general profile, who participate in devising and testing the different metadata subsystems, manage them and assist in-house and external entities in using the system; currently they are 12 technicians with this profile.
- Nomenclaturists, who normally have degrees in economics and study and devise national classifications, monitor EU and international work on studying and devising statistical classifications, assist in-house and external entities in using classifications and give expert

- Terminologists, who have language and literature qualifications and belong to the concepts subsystem. They assist the Production Departments and SC working groups in designing conceptual systems, drafting definitions, arranging for the translation of concepts into English and giving expert opinions; currently the Metadata Unit has 1 technician with this profile.
- Some metadata specialists represent the SP in intra- and extra-community bodies and participate in statistical cooperation programmes.
- Responsible of implementing the SDMX standard: This is a specialist who plays an active part in the Eurostat Task Force to revise the Content Oriented Guidelines. S/he is responsible for studying the SDMX standard so that the metadata system can be adapted to its requirements.

Organisational units and responsibilities: Metadata Unit in the Methodology and Information System Department, Application Development Unit, Production departments

- The main duties of the Metadata Unit are: Design, coordination of development and permanent management of all aspects of the NSS metadata system; Coordination of the technical approval process of Surveys; Management of classifications that are used throughout the NSS.
- The metadata subsystem managers and some classifications managers belong to this unit. There are other classifications managers in the Production Departments.
- The Metadata Unit does not have its own IT specialists. Experts from the Application Development Unit and also the Methodology and Information System Department provide IT services.

Slovenia

Staff roles and responsibilities:

- No permanent metadata management team right now
- General Methodology and Standards Unit: classification server, registry of surveys, survey instances, annual programme of statistical surveys, quality reports, terminology database, advance calendar release, questionnaires and statistical processes
- EDP Infrastructure and Technology Unit: SDMX and archiving
- Department for Electronic Dissemination: metadata for dissemination

South Africa

Staff roles and responsibilities: System Developers, Data Quality Officers and Specialists, Methodology and Standards Professionals, IT help desk technicians

- System developers: Stats SA developers should be knowledgeable to maintain, upgrade and/or enhance the system.
- Data Quality Officers and Specialists: The Data Quality Officers and Specialists are trained on how to use the system. They are also trained to be trainers (“train the trainer”).
- Methodology and Standards Professionals: The Methodology and Standards staff members provide support by developing Policies and Standards. They are subject matter experts in survey operations. They are also involved during the design phase in order to help explain and clarify the requirements.

Organisational units and responsibilities: Methodology and Standards division, survey divisions, IT department

- The network infrastructure for both development and user environments is supported by the IT department. This includes configuring the environments as well as housing the different servers in the data centre of the organization. The databases are also managed by the IT department. Outsourcing is done in conjunction with the IT department, who are in the process of raising their skill level.

Development project: ESDMF

- Project Managers: The Stats SA project manager works closely with the supplier’s project manager. They bridge the gaps between the two organizations and make sure that the deliverables are managed properly and on time.

- The IT procurement group was used to procure all the hardware and software used in the development and deployment of the system.
- In order to understand the user requirements, we engaged the survey divisions as pilot groups Staff roles and responsibilities: IT specialists, metadata system manager, metadata subsystem.
- The Survey Metadata Capture Tool can be used by different users depending on the roles that they were assigned. For example, a Capturer could capture metadata but this must be approved by an Approver, who is usually the supervisor or manager. There is also a role of viewer, whereby metadata could be viewed but the rights are restricted. For example, a viewer cannot edit, change or approve metadata.

Sweden (updated)

- Decentralized responsibility for registering metadata except for statistical databases for which metadata is registered centrally
- Process department (PCA) responsible for functionality of the production process including IT-tools, methods, documentation, support and training
    - process managers are responsible for the business process model and operation of the subprocesses
    - KMI (classifications, metadata and content harmonization) is responsible for harmonization, training and description of statistics
- Research and Development Department is responsible for integrating metadata in overall statistical system (enterprise architects); also includes quality management system
- Subject-matter departments responsible for documenting the statistics in SCBOK, and MetaPlus
    - each department has a documentation manager for training and for review of documentation
- Communication Department responsible for website and statistical databases and their metadata

UNIDO

- 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

5.3 Training and knowledge management

Albania (new)

- All INSTAT employees are trained on an overview of metadata concepts and how to fill the PSZDOC template
- Workshop organized by INSTAT with Balkan countries to present the metadata system
- IT developers are trained in MS SQL, Server DBMS and MS VB .NET development tools

Australia

- DMS provides a range of training. This includes an overview of concepts and systems related to metadata management. Such training is regularly made available to new starters within the ABS and other staff
- A Corporate Metadata Repository (CMR) Assistant is available from the home page of the ABS intranet. This provides a portal to overview and detailed information about the available facilities as well as related policies, guidelines and training courses. It also provides direct access to the facilities themselves by allowing users to click on the component of interest as represented in a high level diagram showing how the various facilities fit together.
- The economic and social statistics areas provide training that includes explanations of how the CMR facilities fit within, and are used within, their business processes. Corporate "Assistants" related to Business Statistics, to Household Surveys and to Publishing cross reference relevant content from the CMR Assistant where appropriate.
- The training about dissemination processes in the ABS likewise includes information about how content defined in the CMR can be drawn into the various dissemination

channels and made available outside the ABS.

- DMS provides development assistance and input on the components of these training courses that relate to the CMR. The strategy of presenting information about the CMR in the context of a particular wider business process, rather than trying to present everything about it exhaustively in a major CMR specific training program, appears to be working very well.

Austria (new)

- Two courses on metadata: Definition of Electronic Questionnaires with the e-Quest Metadata Manager software, and Quality Reporting
- As part of the TQM, standard documentation is reviewed by a subgroup of the Statistical Council

Canada

- The Standards Division has been actively reaching out to the author divisions and staff in general to inform about the purpose and the requirements of the IMDB. Some of the activities included three workshops; the addition of a module on the IMDB to the Statistics Canada's flagship course on business statistics and national accounts; several individual work sessions with specific divisions; and formal presentations to committees.

Croatia

- the IT developers should be trained and experienced in database design and development tools
- the IT developers should also have a rather thorough knowledge of statistical activities.
- the users within CBS should be trained in using the metadata maintenance tool to enter and maintain their metadata, to create a new survey upon another one, to handle versions and history etc.
- detailed user manuals, workshops and hands-on trainings are planned for statisticians.
- training will be prepared and performed by CROMETA developers

Czech Republic

- The SMS department is preparing a number of training activities within the preparation of individual SMS subsystems: nomenclatures and classifications, description of variables, information on TASKS subsystem, and a general course on SMS development. They are included in the CZSO's training system.
- Also training in currently used SMS terminology (incl. selected international organisations).
- SMS is also organizing training for subject-matter statisticians and designers of statistical tasks.
- Separate CZSO intranet web pages have been built and regularly updated to comply with the need of documentation and information sharing between project teams of SIS Redesign, SMS and Public Database. Documents approved by the project teams are available to all CZSO staff members. These web pages proved to be an important tool for dissemination of SMS information inside the Office

Finland (new)

- Metadata experts – working in the Classification and Metadata Services Unit are mentored and guided by senior experts; staff participate in ESTP courses in metadata
- Statistical departments – basic and advanced courses for new recruits and statisticians includes content on present metadata systems including classifications, concepts and archive data management
- Developing training on new common metadata systems and tools as part of the new ICT strategy

Germany (new)

- Metadata are covered by courses on the GENESIS cube database
- RDCs run courses for internal users on the RDC metadata system
- in 2009, this work has been passed onto the Destatis metadata unit

Netherlands (new)

- metadata is stored by the owners of the metadata with assistance of well-trained metadata experts

New Zealand

- Teams within our information management group and standards, solutions and

capability group are currently involved in developing our metadata processes and infrastructure. These groups will also be made responsible for the ongoing training, knowledge management and support of the metadata solution.

Norway

- All new employees participate in an eight day course designed to give them an overview of all areas in Statistics Norway. 2 hours of this course is devoted to metadata in theory and in practice.
- All employees can attend the metadata forum, held approximately twice a year, where relevant problems and projects are presented and discussed.
- A senior adviser from the Division for Statistical Methods and Standards has regular meetings with all division leaders to discuss metadata and gives metadata presentations at all three levels of the organization.

Portugal

- The current training plan that will be held every year consists of five courses: Integrated metadata system, Terminology, Classifications systems, Variables subsystem and Methodological documentation. User manuals are being prepared for the training courses.
- The intranet has a glossary of metadata terminology containing the concepts used by the integrated metadata system.
- The statisticians in the Metadata Unit participate in the above courses, EU-level courses and international conferences and practice high-level self-study.

Slovenia (new)

- SORS prepares and delivers targeted seminars and workshops on information systems
- Emphasis on self-education at the workplace

South Africa

- Users are required to spend at least a day in a training session, taking them through the functionality of the system as well as how to use it.
- The Training Manual is used during the training sessions. The Training Manual contains complete descriptions of the system. The users can also use this document for reference purposes.
- The system is designed such that tool tips (online help) are available to the user when hovering over certain areas of the user interface. This allows the user to have information directly at a point of need without having to go through the Training Manual.

Sweden (updated)

- Process department is responsible for training on the metadata systems and templates
- KMI and metadata experts are responsible for training on classifications, metadata and content harmonization – as part of the MetaPlus and other templates
- Documentation network (where documentation managers are members) is responsible for planning the content of training sessions
- Responsibility for classifications is decentralized and so a classification group has been created to exchange experiences and work on the harmonization of variants

UNIDO

- 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.
- no other special training for the staff was necessary since all statisticians participated actively in the specification and the development of the system

5.4 Partnerships and cooperation

Albania (new)

- metadata project is financed by the Swedish International Development Cooperation Agency and Statistics Sweden – includes study visits, missions from international experts in metadata concepts
- INSTAT uses SCBDoc template and MetaPlus applications
- INSTAT participates in METIS and with the OECD

Australia

- very keen to share information and experiences and to collaborate within METIS generally, as well as on a narrower (eg bilateral or "working group") basis
- a second major international opportunity for partnership and cooperation is seen to be

	around SDMX • contributes actively to international committees associated with other metadata standards of relevance to it, such as ISO 11179. • interest in collaborating on relevant software development, whether on an open source or alternative basis • the National Statistical Service (NSS) provides many opportunities for various collaborations. Many of these collaborations are within Australia but they also include international collaborations, such as with the US Bureau of the Census. • the NSS initiative takes the ABS … into collaborating with other metadata communities, such as the geospatial community, the research community, and others. • a recent collaborative project, for example, with a state government agency and the university sector involved developing "injectors" for technical metadata about usage rights under the Creative Commons framework
Austria (new)	• some metadata projects were conducted with the University of Vienna • international conferences and work sessions – METIS, Metadata Working Group
Canada	• under the UNECE's METIS steering committee, a Common Metadata Framework is being developed, which is used by the Agency to provide guidance. • uses DDI but to support microdata files and is mapping the IMDB metadata standard to DDI
Croatia	• cooperation with Swedish International Development Cooperation Agency (Sida) and Statistics Sweden (SCB) … included consultations carried on through numerous missions, study visits, IT education, procurement of hardware, software and literature etc
Czech Republic	• the SMS is an integral part of the State Statistical Service provided by the CZSO. The ongoing "item inventory and analysis" is coordinated with the workplaces of the statistical service in central government institutions.
Finland (new)	• Statistics Finland participates in working groups for standard classifications and standards at the national and international level • Attend METIS meetings • member of PC-Axis Reference Group • spatial metadata experts follow INSPIRE work and attend meetings
Germany (new)	• Destatis participates in METIS meetings • Knowledge sharing with other statistical offices – i.e., Swiss Federal Statistical Office (Neuchâtel implementation), Statistics Norway (redesign of classification databases in Germany and Norway) • based on experience of the decentralized metadata repositories, which were not coherent in format, data model and terminology, it was realized that a shared metadata model and common terminology was needed
Netherlands (new)	• none
New Zealand	• no text provided
Norway	• participates in relevant international meetings e.g. METIS and informal groups e.g. Neuchâtel (classifications and variables) • strong tradition of Scandinavian collaboration (e.g. dissemination database StatBank) and collaboration within the Statistical Open Source group (e.g. architecture)
Portugal	• Statistics Canada's IMDB project was the main source of reference in developing the SP's integrated metadata system. When SP first began developing the system in 2003, some of its members visited Statistics Canada, where they followed a three-day programme set up by that agency • the Statistical Office of the Republic of Slovenia contacted the SP with a view to learning about its variables subsystem in detail… the SP provided the Slovenian agency with the subsystem's data model and has also responded to later requests.

- co-ordination of economic classifications as part of a statistical cooperation project with the 5 Portuguese-speaking African countries. Consultancy services have also been provided to a project to develop a common integrated economic nomenclature system for the five countries.
- SP also belongs to the Eurostat Metadata Task Force, which is analysing the main components of the SDMX Content-Oriented Guidelines

Slovenia (new)

- Participates in relevant international meetings – METIS, OECD and Eurostat meetings
- Member of PC-Axis reference group
- Classification server was developed after the SNZ model
- Projects since 1997 has been financed within PHARE and 2005 Transition Facility Program

South Africa

- In Latvia, we learned that during the development of their system, their outsourced supplier took a while to understand the business of the statistical organization. We also learned the importance of having a solid foundation in the definition of metadata (e.g., Bo Sundgren's model).
- In Ireland, we learned about the issues regarding communication between the customer and the supplier. Additionally, they had the same problem as in Latvia in that the development of their system also took longer than originally planned.
- In Slovenia, their metadata model is also based on Bo Sundgren's model
- From New Zealand, we adopted a few of their practices. One of their experts helped us to evaluate the respondents to the tender for the development of the ESDMF.
- In our trip to Australia, we learned that in order to have a successful data warehouse project, there is a need to develop policies and standards which will define how the system should be designed.
- Experts from Sweden occasionally came to Stats SA to advise us on various aspects of metadata and statistical production processes
- Last year (2006), we met Alice Born (from Stats Canada) when we attended the METIS conference. We engaged her regarding their development efforts of their metadata system, Integrated Metadata Data Base (IMDB). Consultants from Canada help us in other projects within Stats SA. During their tenure we engage them for advice and other consultation.
- We used the Corporate Metadata Repository (CMR) model by Dan Gillman, from the US Bureau of Statistics in our understanding the metadata model, especially with regard to the ISO 11179 Specification. We also sent our metadata model to him and other metadata experts for review and critique

Sweden (updated)

- Statistics Sweden is a member of the Neuchâtel group and the PC-Axis Nordic co-operation

UNIDO

- no text provided

IV. LESSONS LEARNED

Albania (new) no text provided

Australia

1. While technology is a vital enabler, metadata management should be driven, governed and presented as primarily a business issue rather than a technical issue.

- This requires proponents of metadata management focus first on business outcomes and benefits (e.g., improved productivity, increased utility of statistical outputs) rather than on metadata management itself.
- Metadata management as a topic in its own right is of interest to very few, and is typically viewed as a technical specialisation. Its potential as one (of a number of) means to achieve business process improvement is generally acknowledged. A key challenge is to demonstrate, in practical terms meaningful to business areas, that it should be one of the preferred means - and one that is supported through investment and through business practices and culture. Within reason, the less the term "metadata"

and the names of various metadata standards are used in discussions with senior management and business areas the better - the focus should be on what will be different, and what outcomes will be achieved, from a business perspective.

- "Metadata projects" should not be designed and promoted as "IT system developments" but rather focus on the development and deployment of new and improved capabilities, business processes etc. Such projects will often include new or extended IT systems but they should not be "about" IT systems. Among other drawbacks, narrowing the focus to IT systems will mean business areas - at best - see themselves as relatively remote stakeholders with some interest in the results of the project rather than feeling they are active participants with direct roles to play in ensuring the success of the project.

2. All high level organisational units need to be engaged by the metadata management program and have defined responsibilities in relation to it.

- Some units' primary responsibilities may simply be to contribute to corporate sign off on the objectives, strategies, policies and high level design of deliverables (systems and processes) and then to acceptance test, take up and apply the outputs in an agreed manner to contribute to the achievement of the corporate outcomes sought from the project.
- Other units will have a much more extensive role in terms of leadership, co-ordination, business analysis, design, development, implementation and ongoing management of systems and processes.
- If only a few specific organisational units are seen to have a direct stake in the project then it's much less likely to achieve overall success.

3. It's become more and more apparent over time that applying externally recognised and supported standards, in regard to design of data models for example, has a lot of benefits - including as a means of building upon a wealth of intellectual efforts and experiences from others.

- At the same time, application of standards must be driven, and moderated, by the organisation's particular context and needs. The underlying effectiveness of the infrastructure should not be sacrificed in favour of complying "to the letter" with a standard, although the business case and the management arrangements for any divergence need to be defined and agreed.

4. In addition to developing and deploying infrastructure, a metadata management project should be understood, and managed, as a "cultural change" initiative for an organisation. Metadata management aims to make information explicit, visible and reusable (in whole or in part) - with these aims requiring a somewhat standardised and structured approach. This can be a "culture shock" for some business areas who are used to operating in a more autonomous and self contained and often a less structured manner.

- It needs to be acknowledged there were sometimes benefits from the former approach and there will be some overheads with the changed approach. If at all possible, however, it needs to be demonstrated - or at least plausibly posited - there will be net positive benefits in practice (not just notionally) from the changed approach even if, eg, many of those benefits accrue over time - rather than being immediate - and/or accrue to users of the content rather than producers of the content.
- Sharing and re-use can lead to concerns about loss of absolute "control" over metadata. It is important to ensure practical processes and governance around content use and change management (eg stakeholder consultation, ease of resolution/divergence if a required content change isn't tenable for one of the users of the existing content) address legitimate concerns in this regard
- Note also the cultural change aspects discussed in Section 5.5 in terms of moving from "local optimisation" to a paradigm of "global optimisation".

5. Sufficient attention needs to be focused, by the project team and by other areas, on ensuring the metadata management infrastructure (systems and processes) is fully integrated with other business processes and IT infrastructure rather than being a "stand alone" development.

- This needs to factored in from high level design onwards. This, in turn, requires that as part of the initial requirements gathering, analysis and sign off phase there is detailed attention focused on practical matters related to implementation, including uptake and

ongoing use as part of end to end business processes.

- This is also a reason why acceptance testing of deliverables by a fully representative selection of the business areas expected to use them is essential. The aim of this testing is not so much to confirm the specifications have been implemented faithfully (detailed system testing should already have been completed) but that the results meet practical business needs, including integrating with other workflows and systems and meeting performance and other usability requirements.
- "Acceptance" testing should mean just that. If (for whatever reason) what has been delivered is not yet at a stage where it is fundamentally fit for purpose from a business perspective then it should not be deployed in its current form. (On the other hand, if the deliverable is imperfect but basically "fit for purpose" then the remaining issues may be held over to be addressed in a later release.) The phase ends either with business agreement the deliverable is fit for use within the broader production environment - possibly with some caveats - or else no release occurs.
- Sound project management and engagement with business stakeholders in earlier phases should minimise the risk of failure at the Acceptance Testing stage. That said, it is counterproductive for all concerned if software that is not fit for purpose is forced on business areas.

6. In addition to allowing sufficient time and resources for the business analysis, design and development process it is crucial there is sufficient resourcing focused on

- implementation of the new infrastructure
- includes training, best practice advice and technical troubleshooting support for business users
- maintaining and upgrading the infrastructure as business requirements, and as other elements of the IT environment, evolve over time
- co-ordinating and promoting "outcome realisation" from the investment.

7. Business areas must be able to engage with implementation processes.

- In many cases there may need to be scope for business areas to negotiate and agree (not decide unilaterally) short term or longer term exemptions from, or variations on, the standard implementation process.
- Exemptions and variations should be actively managed and reviewed with the aim of achieving convergence over time wherever practicable
- Metadata systems should clearly identify preferred and non preferred definitions and structures, so that - wherever practicable - areas with a need to diverge from standard practices and definitions remain "within the system" while at the same time those practices/definitions are clearly identified as non preferred.
- Feedback from business areas needs to be able to influence the details of the implementation process. For example, if it appears too many exemptions and variations will be required it may be that the design of the implementation process doesn't properly reflect business needs and realities.
- If business areas are not provided with a genuine opportunity to "work with" a change process they are more likely to covertly "work around" that process in a manner which undermines the business objectives of the change.

8. Metadata management is largely about connections of various forms, such as

- between documentation of agreed processes, methodologies, definitions and structures and what happens systematically
- between producer and consumer perspectives on statistics
- similarities and differences between different sets of data, different structures and definitions etc

9. Due to the wide variety of roles it must perform, and perspectives it must support, there is not one particular structure/format for metadata that is, in itself, ideal for all purposes.

- The key appears to be modelling and managing metadata in way that can support the different views, and preserve the integrity of the connections underlying these different views.
- The ideal appears to be a relatively simple, robust, standards aligned but highly extensible core model, together with well defined and managed means to map and

transform locally required metadata into and out of that core model and, where necessary, to define, manage and integrate local specialised extensions to that common core.

- A single central metadata model that aims to span all content for all purposes is likely to be too complex, too unwieldy and too static.

10. "Statistical metadata management" is increasingly expected to interoperate with metadata management as practised in other communities (eg geospatial, academic/research) and sectors (eg use of XBRL by businesses and by regulatory agencies). This provides a huge opportunity (as well as a challenge) in being able to efficiently and effectively open up and harness (from statistical and other perspectives) a vastly increased suite of information resources. It also provides a practical affirmation that other communities and sectors recognise the value of metadata and standards, although because their primary purposes vary the details of their schemas and standards also vary.

- This reinforces the previous point. It appears both impractical and undesirable to establish a single approach that supports the primary purpose of each different community and sector. On the other hand, statistical agencies are strategically placed to provide a simple core that might be used to bring together information across communities/sectors and to exchange it across them.

- It also reinforces the value of international standards and collaborations. Most of the community and sector specific standards are internationally based. Rather than each NSO needing to work out mappings "from scratch" there is a lot of opportunity to share a core of analysis and mapping between NSOs.

Austria (new)

It is not a new discovery that the subject of statistical metadata is an extremely complex one. Even now, different individuals may still mean quite different things or place emphasis on different aspects when speaking of metadata. This phenomenon is even more pronounced when these persons stem from different areas of expertise: senior management, subject matter statisticians, methods specialists, IT experts etc.

Papers prepared for the METIS work sessions and the Common Metadata Framework (especially the case studies) have proven very useful, as they provide arguments for discussions with statisticians and top management.

The creation of an integrated system consisting of more than isolated solutions is difficult when there is no organizational unit the main responsibility of which is to deal with the subject of metadata and its usefulness for the NSI - and which is also granted the requisite authority and enjoys the support of top management, so that it can achieve the introduction of integrated and centralized metadata systems even against the possible resistance of subject matter departments.

That metadata projects are best carried out using an interdisciplinary approach (and not as IT projects) has long been recognized in expert circles. In practice, however, it appears that the qualified subject matter statisticians continually suffer from such a heavy workload that they have no time to spare for complicated conceptual work (e.g., Statistics Austria has reduced the number of personnel by about a third since its separation from the federal civil service in the year 2000).

Many statisticians associate the concept of "metadata" with the notion of "additional work" (which for instance actually was the case when the standard documentations were introduced). This leads them to resist new metadata systems.

The idea of developing specialized tools for editing, administrating and (re-)using metadata with an end-to-end approach regarding the statistical life cycle often encounters resistance among statisticians because the introduction of such tools will result in changes to work processes which they have been familiar with for many years.

Statistics, however, is not the only field of activity in which the creation and usage of

metadata can be seen as part of the job description. In order to produce software of high quality and in an economic way, the availability of tools - to support the management of "software metadata" (including the source code of the programs) and to provide services to alleviate the software engineers' work - has long been recognized as necessary. Especially when several programmers are cooperating in a software project, the storage and administration of all information items in a central repository seems indispensable.

The production of statistics exhibits a high degree of similarity to the production of software. However, in statistics the advantages offered by specialized tools and a centralized metadata repository are not yet generally accepted.

For the development of systems for the collection and administration of passive metadata, the cost factor presents a particular obstacle. Passive metadata are an integral component of statistical information. Their availability and easy accessibility contribute to the quality of statistical products, but in many cases do not result in cost reductions (they may even increase the work load of subject matter statisticians). Opportunity costs caused by the non-existence of centralized end-to-end metadata systems are rarely found in accounting systems. Thus high investments are accompanied "only" by a gradual gain in quality (which may not even be recognized by all user groups). Under these circumstances it is understandable that in times of economic crisis the willingness to invest in metadata projects is not high.

The concept of "high-quality statistics" is a dynamic one. The needs and requirements of users are changing and will probably increase in the future, e.g. with regard to harmonization of statistics or the linkage of data with relevant metadata items (respectively linkage of metadata items with related metadata items), so that they can be accessed at the push of a button. If metadata are stored in the continuous text of bulky documents, these new requirements cannot be met. The management of metadata in an "atomic" and structured form, however, is a challenge with respect to both financial resources and personnel.

The fundamental principles of metadata management, which have been defined by experts during recent years (and which can be found, for example, in part A of the Common Metadata Framework) will become more and more commonly accepted standards and state of the art for the production and dissemination of statistical information.
The task of implementing these standards can certainly not be carried out at short notice. In this respect, it is not easy to answer the question whether to continue building isolated metadata systems whenever the need for one specific system arises, or whether to strive for an integrated system based on a global architecture. The first approach is certainly less expensive in the short run and produces quicker results, but in the long term it will cause quite substantial "repair" costs.

What metadata should actually be collected for and provided to external and internal users, and in what form should they be provided? This is a fundamental question on which opinions within Statistics Austria are divided. The search for an answer should not be postponed just because it is clear from the start that up-to-date solutions will require high investments in time and money. The answer should rather be given as soon as possible in order to ensure from the start that the solutions - which must be planned and implemented step-by-step in accordance with budgetary constraints and on a long-term time scale - will be built to last.

Canada

The progress made during the variable documentation phase, as well as with the methodology and data accuracy documentation phase, leads us to conclude that it is more efficient to start documenting the metadata right at the outset of any new survey design or redesign, and at the data collection stage of surveys, instead of documenting them only after they have released data, which was our practice in the past.

All of these plans and priorities for the further development and use of the IMDB critically depends on the quality of the metadata it contains, which in turn depends on the co-operation, motivation and competencies of metadata authors. For a metadata system to be successful, communications and training activities need to be continuously implemented so that authors are fully aware and convinced of the importance of metadata and their role in its creation and maintenance. They need to know and understand the specific requirements of the IMDB and integrate the creation and updating of metadata in their routine business process. Continued management support and attention is necessary but, in addition, authors need to know the how and why of metadata.

Controlled deployment of the updating system to the desktops of motivated and trained divisions, coupled with information sessions for all data producing divisions, incorporation of metadata training in the corporate flagship training programs, and workshops on metadata and providing additional information on plans and progress on the IMDB project are all activities that have improved the level of awareness and understanding regarding the use and importance of metadata at Statistics Canada.

Croatia

The most important questions are still unanswered since the central metadata management system is not deployed yet. The complete ISIS information system in general and CROMETA system in particular will force big changes upon the overall culture of CBS. The degree of content or discontent by the majority of statisticians when they start using the metadata maintenance tool is yet to be learned. In any case, we expect resistance from a number of subject-matter experts, especially those who cherish very much the legacy from past times when it was usual practice to order a tailor-made data processing system from the IT department.

Therefore, the 'Lessons learned' here apply mainly to software development activities and this is by far less important than the overall cultural changes that will be met by the deployment of the new ISIS.

1. The most important lesson learned is that there is no serious development when there is no development team appointed to this and only this project. This applies to IT developers as well as statisticians. Of course we knew that even before we started the project but we could not afford to have experts unavailable to regular production for a longer period of time. So we entered a vicious circle: we wanted to develop software to make production easier but we could not develop because we had to handle the production. For this reason the development lasted much longer than planned.
2. The support from top management is crucial.
3. Teamwork is very important and now it is enhanced with practical tools such as SharePoint etc.
4. It is obvious from experiences of other NSOs that the involvement of statisticians is crucial for the project. Therefore we tried from the beginning to include selected statisticians in the development through various forms of cooperation but somehow it always ended after two or three meetings (see item1.).
5. Strict project management. It is the responsibility of the project management that statisticians fell out of development activities sooner or later.
6. The most painful lesson learned was that there is no project interesting and challenging enough to keep young and well educated IT experts from going to better paid jobs. IT experts in government bodies are paid two or three times less than in private sector and this needs no further comment. The CROMETA project started with 16 people (more or less involved) and ended with 3.

Czech Republic

Some most important experiences and conclusions from our practice:

- SMS strategy in terms of contents and methodology must be fully in the responsibility of the statistical office,
- SMS design and implementation should be organized in the multidisciplinary working teams;

- design and implementation of the SMS project must be managed and systematically monitored by the top management,
- it is necessary to persistently obey the SMS system principles and to maintain a positive motivation of most wide circle of subject-matter statisticians and methodologists; in this respect the CZSO benefited from involvement of external expert as a consultant to the Office President,
- consistent co-ordination of time-scheduled workloads in the SMS project, the SIS Redesign project and current activities of the Office,
- purchasing of financial funds must be systematically monitored by the statistical office in relation to the stage of the project implementation, on the basis of functional specification and qualified estimate of man-hours. It is important to use all potential sources of funding (external and internal sources),
- financial costs of the operational running of the SMS should be covered from the Office budget.

Finland (new)

The basis for metadata work must be seen to reside in the contents, not in the technology. Commitment by the Management and their support to the work is crucial for the statistical units to be able to provide the contribution needed to the development work and for ensuring that the work will be sustained.

The centralised metadata system should support the harmonisation of the production of statistics to a sufficient degree, thus making it more effective, but it should also be flexible enough to a certain extent to serve what statistics specifically call for.

Germany (new)

- Metadata management is a communication challenge. We found two issues were particularly difficult to communicate:
- Metadata management is tricky. Statistical data is inherently volatile. For any given data, an endless number of transformations are possible producing an endless amount of metadata. With the distribution of modern IT-systems there is hardly any limit to producing endless variations of the same dataset.
- Metadata can be more than just documentation. The same information that is used to transform (produce) data can be used to document it and vice versa.
- As we are faced with multiple stakeholders, several isolated decisions taken by governing committees and a variety of IT-systems in place, it would probably be useful to develop a metadata strategy. Such a strategy might help the organisation to focus on important projects and provide a coordinated approach ensuring that systems are able to interact. Distributing the energy of an organisation across too many unrelated tasks easily drains away resources without delivering satisfactory results. Drafting such a strategy, however, also consumes resources and requires a deeper understanding of the problems.
- The advantages and disadvantages of a metadata model can often only be properly evaluated once an IT-system is in place. It is therefore important to learn from evaluations of existing systems.
- Considerable effort went into formulating metadata models. Having evaluated some of them, we feel that the existing models do bear some similarities. A perfect model may not exist, especially since the resulting implementation usually involves some compromise. No database can be endlessly complex. But on a more conceptual level there seems to be some convergence. Indeed there might even be a structure inherent to the metadata of (official) statistics. Thus, the quest for the "real" metadata model might be less a matter of design than of discovery.
- In a federal system, national coordination usually requires a lot of resources from all partners in the system. Understandably, international cooperation is then often seen as being of lesser importance. Despite this, international cooperation has substantially helped the metadata team at Destatis to understand the subject of metadata management. The development of IT-systems consumes a lot of resources. We feel it helps to build on existing international knowledge and that it minimises risks and

maximises return on investment.

Netherlands (new)

- Small projects that deliver in short cycles;
- Use of external off-the-shelf software is possible without too much adjustments in specs;
- Keep in control of outsourced development activities;
- It is a challenge to formulate a convincing business case for metadata;
- Develop a metadata architecture in order to direct the development of metadata models that will be needed as new features for the storage, retrieval and transformation of statistical data will arise.

New Zealand

1. Apart from 'basic' principles, metadata principles are quite difficult. To get a good understanding of and this makes communication of them even harder. As it is extremely important to have organisational buy-in, the communication of the organisation metadata principles and associated model is something that needs some strong consideration.
2. Every-one has a view on what metadata they need - the list of metadata requirements / elements can be endless. Given the breadth of metadata - an incremental approach to the delivery of storage facilities is fundamental.
3. Establish a metadata framework upon which discussions can be based that best fits your organisation - we have agreed on MetaNet, supplemented with SDMX. As Statisticians we love frameworks so having one makes life a lot easier. You could argue that the framework is irrelevant but its the common language you aim to use.
4. There is a need to consider the audience of the metadata. The table about users covers some of this, but there is also the model where some basic metadata is supplied (e.g. Dublin Core) that will meet one need but this will then be further extended to satisfy another need and then extended even further to meet another need.
5. To make data re-use a reality there is a need to go back to 1st principles, i.e. what is the concept behind the data item. Surprisingly it might be difficult for some subject matter areas to identify these 1st principles easily, particularly if the collection has been in existence for some time.
6. Some metadata is better than no metadata - as long as it is of good quality. Our experience around classifications is that there are non-standard classifications used and providing a centralised environment to support these is much better than having an 'black market' running counter to the organisational approach. Once you have the centralised environment with standard & non-standard metadata you are in a much better position to clean-up the non-standard material.
7. Without significant governance it is very easy to start with a generic service concept and yet still deliver a silo solution. The ongoing upgrade of all generic services is needed to avoid this.
8. Expecting delivery of generic services from input / output specific projects leads to significant tensions, particularly in relation to added scope elements within fixed resource schedules. Delivery of business services at the same time as developing and delivering the underlying architecture services adds significant complexity to implementation. The approach with the development of the core infrastructure components within the special project was selected to overcome this problem.
9. The adoption and implementation of SOA as a Statistical Information Architecture requires a significant mind shift from data processing to enabling enterprise business processes through the delivery of enterprise services.
10. Skilled resources, familiar with SOA concepts and application are very difficult to recruit, and equally difficult to grow.
11. The move from 'silo systems' to a BmTS type model is a major challenge that should not be under-estimated.
12. Having an active Standards Governance Committee, made up of senior representatives from across the organisation (ours has the 3 DGSs on it), is a very useful thing to have in place. This forum provides an environment which standards can be discussed & agreed and the Committee can take on the role of the 'authority to answer to' if need be.
13. Well defined relationship between data and metadata is very important, the approach with direct connection between data element defined as statistical fact and metadata

dimensions proved to be successful because we were able to test and utilize the concept before the (costly) development of metadata management systems.
14. Be prepared for survey-specific requirements: the BPM exercise is absolutely needed to define the common processes and identify potentially required survey-specific features.
15. Do not expect to get it 100% right the very first time.

Norway

1. Top management support is essential.
2. Make a metadata strategy. It is important that we can refer to formal documents like the metadata- and IT-strategy (which has been approved by the board of directors) in our metadata work. In the same way it is useful that the list of key metadata terms promoted for use within the statistical office has an official "stamp".
3. Use step-wise development of metadata systems with active user involvement and regular delivery of functionality.
4. Ensure continuous follow-up of progress and quality with direct feedback to users and regular reports to middle and top management. One of the biggest challenges in management of metadata is allocating the necessary resources. Releasing good quality statistics within the planned time schedule is the primary task for the subject matter divisions and documentation will often have a lower priority. It is therefore crucial that the management stresses the importance of documentation and increases the status for this kind of work.
5. Harmonising variables between subject matter divisions is also a considerable challenge and an important tool to improve the quality of metadata. Several subject matter divisions may use the same variable names, but define them differently. In some cases this is necessary because of laws and regulations, but this is not always the case. We have meetings where contact persons from divisions using variables with similar names come together and discuss the definitions, e.g. if a division could change the wording of their definition to such an extent that other divisions might use it as well, which would allow us to reduce the number of definitions to one in stead of e.g. three. This is a time consuming work which we have started, but which will require a lot more of resources, both to monitor where harmonisation is needed and to do the job.
6. The possibility to release metadata on the Internet makes it easier to motivate subject matter divisions to document metadata and improve metadata quality.
7. We think that to really make metadata work a natural part of everyday life in the subject matter divisions, we have to include the metadata systems in the production cycle. Then we can establish routines where the handling of metadata is included in all relevant production steps. So far the metadata work in Statistics Norway has been focused on implementing metadata systems and filling them with relevant documentation. This year we will start investigating the role of metadata(systems) in the production cycle.

Portugal

1. We have certainly learned some lessons from the implementation of the integrated metadata system, which has been more systematic in the last six years, some because we have seen that our options have had a positive effect and others because we have realised the form they should have taken in order to be more successful. We are even making some changes in the formal circuits of some subsystems with a view to greater efficiency and quality in the results obtained.
2. Involvement of the institution's top management was fundamental and the tie-in of the creation of documentation with formal and standardised procedures has been an excellent way of keeping documentation up to date at the SP.
3. Designing a metadata system not only requires considerable knowledge of statistical production, but also means leaving behind some habits acquired in this area. A great capacity for abstraction and tidy, integrated thinking is also necessary. An institution has specialists with all these capabilities but not always with all of them at the same time. The teams chosen to implement these systems must consist of specialists with different profiles among those mentioned, because they complement each other. The IT technicians who develop applications must participate from the start.
4. We believe that it is essential to develop prototype systems before final implementation. Prototyping is the best way to test a system's design, detect strong and

weak points and come up with experience-based alternatives for the weak points. When designing a system like this, it is very hard to give an appropriate description of all its functions without prior experience. Even the workflow of procedures may need some adjustments.

5. Must be given training to statisticians, not only in the use of applications but also, and above all, about the concepts underlying the system and workflow of procedures. The introduction of the position of survey manager has fostered cooperation and dialogue between production, metadata and dissemination. The distribution of terminology associated with each metadata subsystem is having a beneficial effect at the SP as it encourages the use of a language common to all profiles using the system.

6. After the classification subsystem was made available to the general public, we began to receive some complaints about its usability and decided to conduct some usability tests. The test results showed us the difficulties that people experienced when using the system and we decided to redo some of the navigation in the consultation application. When we implement the methodological documentation subsystem, scheduled on the beginning of 2008, we have decided to conduct usability tests in the prototype phase of the consultation and publication application so that we do not need to redo any parts of the system after it goes into production.

Slovenia (new)

Participation in pilot projects enables less experienced employees to gain the experience necessary for independent work. Due to the similarity of the statistical process, it is very important for the IT personnel in SORS to gain experience from other statistical offices, which is ensured through participation in expert conferences and bilateral cooperation with foreign offices.

Metadata usage in production tools has shown many opportunities but also a lot of new challenges.

In 2009 within total quality management priorities will cover:
- activities to promote the cooperation of enterprises in data reporting;
- concern for the consistent monitoring and analysis of response burden;
- implementation of the process of standardising questionnaires in statistical surveys;
- preparation of documentation on the consistency and coherence of statistical data from various surveys;
- preparation of internal rules and procedures on revisions of published statistical data;
- establishing internal rules and procedures for keeping documentation for the time series breaks.

Great attention will be focused on the preparation of internal methodological manuals (i.e. textbooks, handbooks, presentations, etc.) for individual parts of the statistical process.

By providing the necessary technical and expert support for establishing statistical methodology, SORS attempts to increase interest in implementing the quality standards of the European Statistical System among authorised producers of national statistics. This will cover also the formulation of methodological explanations and a release calendar. At the same time a short course is being planned on imputing missing values, data editing in statistical surveys, on sampling methods and data dissemination.

SORS has been building a new metadata driven system during the last few years. This information system is designed in a highly flexible and scalable way, based on modern technological approach (SOA), to assure easy adaptation to future changes, reusability of developed services and further development of the SOA end electronic business at SORS. Relevant actors (users) were engaged in testing in early phases of project to know the new system and to be able to start use it after the end of the project in the most efficient way.
Descriptions of the problems

- Complexity of the project / relatively short time for implementation
- The new project has been very complex and ambitious. Information system covers

metadata management system, entire process of flexible and metadata driven statistical surveying, sophisticated electronic reporting system, Statistical business register and special respondent's management system, seasoned with heterogeneous technological platform and all kinds of integration demands with specific statistic tools.

- User specifications. Consultant did prepare technical documentation during analysis and design phases which has described the future information system; however, the technical documentation in most cases was not transparent enough for users to understand the designed system in details before it was built. The main problem was »translation« from »statistical language« to »IT language« and it caused many misunderstandings.
- Organizational adaptation of SORS. Information system like the newly developed, inevitable results in fundamental adaptation of organizations. It should enable SORS to turn paper-centric statistical surveying into paperless statistical surveying - it is hard to imagine a bigger change. The organisation of data process has been under process of adaptation to implement the metadata driven system successfully in its processes. Implementation of electronic reporting alone demands entirely different "service and customer-centric" organization of data collection process.
- Organization of user testing and bug repair. User testing and bug repair process on some modules was not organized efficiently enough and it did not enable to complete the user testing and bug repair phase timely.
- Some recommendation for the future. We highly recommend to perform smaller projects with strong project leaderships; even smaller projects need two phases: the first is the preparation of business redesign (could be paper-and-pen, no prototypes yet) and the second is the translation of redesign into IT language. The first phase is time consumming and all the details of the data process should be described and explained.

South Africa

The supplier had a difficult time understanding the business of Stats SA, which is statistical production processes. Additionally, the goal of the project is to improve quality, which will help support the vision of Stats SA "to be the preferred supplier of quality statistics". Even in the face of this vision, the supplier failed to recognize that quality was a primary business objective.

Under pressure of meeting the deliverables, the supplier ignored the Skills Transfer Plan, with the result that the Stats SA developers were not involved in the final design and development of the system.

For a project of this magnitude (three years), we decided to break down the deliverables into twelve phases. Each phase was planned to be three months long in duration. Also, each phase was planned to be a complete deliverable in its own right, even though the next phase was planned to build on the previous phases. The first phase was delivered late mainly due to the lack of understanding that the supplier demonstrated. The key is that clear understanding of the requirements is very important in meeting the deliverables as well as milestones for those deliverables.

Sweden (updated)

When building metadata systems:

- Involve different types users at an early stage, we used different groups for:
    - Methodologists
    - Subject matter statisticians
    - System developers
- Make it a content, not an IT -driven project
- Make simple prototypes as early as possible to get input from users, written requirements are to abstract

Be aware of that it is a complicated task and takes a lot of time to get things right.

UNIDO

no text provided