

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

**PART**  
**A**  
Common Metadata Framework

**Statistical Metadata  
in a Corporate Context:**  
A guide for managers



**UNITED NATIONS**  
Geneva, 2009

## **NOTE**

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontier or boundaries.

ECE/CES/1

## Foreword

The United Nations Economic Commission for Europe (UNECE) is pleased to present this publication, Part A of the Common Metadata Framework. The development of a common framework for statistical metadata was initiated by the 2004 Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), which concluded that a framework was needed to provide links and context to previous and current international metadata standards initiatives together with comparisons of selected examples of current practices, to help statistical organizations to develop their metadata systems.

The Common Metadata Framework is being developed and maintained through discussion and inputs to METIS Work Sessions and Workshops. At the time of writing, it comprises four parts:

- Part A: Statistical Metadata in a Corporate Context. This part highlights the role of metadata systems in statistical organizations and focuses on managerial issues relevant to the corporate governance of statistical metadata systems.
- Part B: Metadata Concepts, Standards, Models and Registries. This part provides information about relevant concepts, international standards and models, and explores the links between them.
- Part C: Metadata and the Statistical Business Process. This part contains explanations, good practices and other material to assist statistical organizations to design and develop a statistical information system relevant to their business requirements.
- Part D: Implementation. This part focuses on the experiences of statistical organizations that have recently implemented or re-engineered their statistical meta-information systems. It consists of a set of case studies published on the METIS-wiki.

The latest versions of the different parts of the Common Metadata Framework, as well as other information on statistical metadata, are available online at [www.unece.org/stats/cmf](http://www.unece.org/stats/cmf). Part A was first published online in 2006. It has been revised and updated during 2008 and 2009 in consultation with the METIS Steering Group.

The underlying purpose of Part A is metadata advocacy, i.e. to raise awareness of the importance of metadata throughout the statistical production process. It also aims to demonstrate that a statistical metadata system is not just an information technology project. It impacts of all areas of statistical production so requires inputs from many different disciplines, as well as strong corporate governance with the active involvement of senior managers.

## Acknowledgements

This publication has been prepared thanks to the collective input of member countries of the Conference of European Statisticians, the Statistical Metadata (METIS) Steering Group and the coordination efforts of UNECE secretariat staff and an independent consultant, Jana Melíšková. Feedback and suggestions for improvement are encouraged and can be sent to the UNECE secretariat: ([support.stat@unece.org](mailto:support.stat@unece.org)).

## TABLE OF CONTENTS

1.	Introduction	5
2.	The role of a statistical metadata system	7
3.	Users of the statistical metadata system	10
4.	Benefits for users	12
5.	Vision, strategy and implementation	19
6.	Core principles for metadata management	32
7.	Corporate governance models for metadata management	33
8.	Case studies and experiences	39
	Australian Bureau of Statistics	39
	Central Bureau of Statistics, Croatia	40
	Czech Statistical Office	41
	German Federal Statistical Office	41
	Statistical Office of the Republic of Slovenia	42
	Statistics Austria	43
	Statistics Canada	45
	Statistics New Zealand	45
	Statistics Norway	46
	Statistics Portugal	47
	Statistics Sweden	48
	Glossary of terms and abbreviations	49
	References	50

## TABLE OF FIGURES

1.	Metadata users	10
2.	Schematic view of the SMS vision and its components	21
3.	Example of metadata objects for statistical production processes	22
4.	Model for integrated management of SMS	26
5.	Perceived challenges to introducing or using SMS	34
6.	Perceived barriers to the provision of effective metadata	35

## 1. Introduction

Statistical metadata systems play a fundamental role in statistical organizations. Such systems comprise the people, processes and technology used to manage statistical metadata. This publication provides guidance to senior managers in understanding the importance of metadata and the systems needed to manage them effectively. It outlines the main issues to consider when establishing and maintaining statistical metadata systems, and how to manage metadata across a statistical organization. It focuses on managerial issues, such as corporate governance of metadata projects and the business benefits of good metadata management.

This publication is Part A of the Common Metadata Framework, an initiative to help statistical organizations to choose the right standards, models and approaches in developing their metadata systems. The Common Metadata Framework is a living repository of knowledge and good practices related to statistical metadata, the latest version is available online at [www.unece.org/stats/cmef](http://www.unece.org/stats/cmef).

A statistical metadata system (SMS) is an important tool for ensuring the goals of the statistical information system are met. Since metadata users have diverse needs, effective management of statistical metadata is strategically important for any statistical organization. The foundation of an effective SMS is to identify metadata users and understand their needs. Experience also shows that to be successful, an SMS must be an integral part of a statistical organization's strategic direction.

Management of an SMS project is a demanding task. Statistical metadata is a developing field, attracting researchers and experts from both inside and outside statistical organizations. International organizations are working together to provide standards and guidance and several projects dealing with different aspects of statistical metadata management have been conducted. Standards and guidelines for statistical metadata have been developed, and are already applied in practice by national and international statistical organizations. These include the Statistical Data and Metadata eXchange (SDMX) initiative, including Content-Oriented Guidelines, the OECD Data and Metadata Reporting and Presentation Handbook, the Generic Statistical Business Process Model and the UNECE Guidelines for Statistical Metadata on the Internet. Work continues to progress through the regular meetings of the joint UNECE / Eurostat / OECD work sessions on statistical metadata.

The experiences of national and international statistical organizations have shown that direct involvement of senior management is a prerequisite for a successfully functioning SMS. It is not sufficient to engage only metadata experts and information technology specialists in this work. The important role that methodologists and subject-matter statisticians play in managing metadata must be recognized. The variety of metadata users, and the cross-cutting nature of metadata related activities throughout the statistical business process, calls for the involvement of top management in metadata projects.

In the past, the main, and sometimes only, role of metadata in a statistical organization was to support the production of official statistics. However, the SMS should go beyond this function to address other requirements. It should be a tool to facilitate the efficient functioning and further development of the whole statistical information system. This requires corporate commitment and systematic management of activities related to design, implementation, maintenance, use, and evaluation of the SMS.

### Target audience

The aim of this publication is to help statistical organizations improve the effectiveness of statistical metadata across all phases of the statistical business process. It addresses the information needs of senior managers in statistical organizations and is intended as a tool to assist statistical metadata experts to develop business cases for a new or enhanced SMS.

Managers, designers, subject-matter specialists, methodologists, information technology experts, and researchers are all involved in SMS projects. They are responsible for different aspects of the SMS, but need a common understanding of the role and complexity of the system. This not only fosters a culture of

teamwork, but ensures clear and consistent communication of the management issues in the SMS business case. This publication provides an SMS project team with a common understanding of statistical metadata in a corporate context: the strategic importance of an SMS, and in particular, the need to integrate the SMS into the overall strategy of the organization.

### What is in this publication?

This publication explores the following topics related to statistical metadata management:

- **Value proposition for metadata management** defines the role and functions of the SMS for statistical organizations. It describes major users of statistical metadata and the benefits an SMS provides to each of them.
- **Metadata management strategies and policy framework** covers the management and preparation of a corporate SMS vision. It presents a brief description of metadata and explains the role they play in statistical business processes. Recommendations for the preparation of a strategic plan for managing metadata are provided.
- **Principles for metadata management** express the fundamental considerations and recommendations for managing metadata. The aim of these principles is to facilitate design, implementation, maintenance and use of a corporate metadata repository.
- **Corporate governance model** provides guidance in good corporate governance of an SMS, taken from the experiences of statistical organizations in the implementation of a metadata management strategy. It explains potential risks and challenges in management, and considers related organizational challenges.
- **Case studies and experiences** include lessons learned in actual implementations and practical experiences in the development of statistical metadata systems.

### How to use this publication?

Readers will get an overview of the **functions** and content of an SMS, the **major steps** in the design and implementation of such a system, potential **users and stakeholders**, and how to prepare or update a **corporate metadata management strategy**.

For hints on how to use the publication when preparing specific parts of a business case, whether it is for a new SMS or enhancement of existing systems, please see the following topics:

**When preparing a value proposition for a corporate system:**

Clarify the role of the SMS in the organization (page 7)

Analyze users' needs and the benefits they will gain from the system (page 10)

**When preparing a vision for statistical metadata management:**

Specify the goals of the SMS (page 19)

Prepare the strategic plan (page 23)

**When preparing management strategies for the SMS** (as an integral part of the vision)

Set up the SMS management strategy across the whole organization (page 25)

Prepare the management strategy spanning the SMS life cycle (page 26)

Follow the core principles for metadata management (page 32)

## 2. The role of a statistical metadata system

### What is a statistical metadata system?

Metadata can be defined as “data that define and describe other data” whereas statistical metadata are “data about statistical data, and comprise data and other documentation that describe objects in a formalised way” (both definitions come from the 2009 edition of the SDMX Metadata Common Vocabulary). Some metadata, however, describe other resources that are relevant to the collection, processing and dissemination of data (for example, questionnaires and publications) rather than relating directly to data themselves, so a slightly wider definition of metadata may need to be considered. With this in mind, the SDMX Metadata Common Vocabulary definition of a Statistical Metadata System (SMS) can be used: “A data processing system that uses, stores and produces statistical metadata”. The term system refers to the people, processes and technology involved in managing statistical metadata.

Metadata have two basic functions. The first is to uniquely and formally define the content and links between objects and processes in the statistical information system. The second is to determine all related technical parameters.

With rapidly developing technologies, it can be challenging to implement efficient strategies for the production and dissemination of statistics. The growing use of the Internet has caused significant changes in the priorities for managing statistical metadata. In the past, priority was often given to technical metadata and information technology challenges, whereas now there has been a clear shift to focusing on content and methodological issues. When designing the SMS, priority should be given to the content of the system. The primary focus should be achieving business outcomes and ensuring the expected benefits can be realised through effective governance and processes.

Due to these changes, transparent and integrated descriptions of information flows within and outside the statistical office are vital. The use of technology for data collection, interactive communication with users, and dissemination of statistics, calls for a coherent and well functioning metadata system.

The implementation of an SMS should be independent of the technology used for statistical data processing. However, the links between the SMS and electronic processing systems must be ensured. Processing of statistical data should be driven by metadata stored in the SMS.

### What is the role of the statistical metadata system?

The success of an SMS can be measured by the extent to which the needs of diverse groups of statistical metadata users are satisfied. The need for metadata is defined by the various activities, tasks and processes carried out in a statistical organization. Therefore, the role of SMS should be understood in the context of these processes and activities.

A basic framework for the role of the SMS in statistical organizations is defined in:

- **The Fundamental Principles of Official Statistics**  
(adopted by the United Nations Statistical Commission in April 1994)  
<http://unstats.un.org/unsd/methods/statorg/FP-English.htm>; and
- **The Principles Governing International Statistical Activities**  
(endorsed by the Committee for Coordination of Statistical Activities in September 2005)  
[http://unstats.un.org/unsd/methods/statorg/Principles\\_stat\\_activities/principles\\_stat\\_activities.htm](http://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/principles_stat_activities.htm).

The principles relevant to metadata management are highlighted below:

### **The Fundamental Principles of Official Statistics**

- Principle 2.** To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.
- Principle 3.** To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.
- Principle 5.** Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.
- Principle 8.** Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.
- Principle 9.** The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

### **The Principles Governing International Statistical Activities**

- Principle 1.** High quality international statistics, accessible for all, are a fundamental element of global information systems.
- Principle 2.** To maintain the trust in international statistics, their production is to be impartial and strictly based on the highest professional standards.
- Principle 4.** Concepts, definitions, classifications, sources, methods and procedures employed in the production of international statistics are chosen to meet professional scientific standards and are made transparent for the users.
- Principle 5.** Sources and methods for data collection are appropriately chosen to ensure timeliness and other aspects of quality, to be cost-efficient and to minimize the reporting burden for data providers.
- Principle 7.** Erroneous interpretation and misuse of statistics are to be immediately appropriately addressed.

The SMS should be a tool enabling a statistical organization to effectively perform the following functions:

1. Planning, designing, implementing and evaluating statistical production processes.
2. Managing, unifying and standardizing workflows and processes.
3. Documenting data collection, storage, evaluation and dissemination.
4. Managing methodological activities, standardizing and documenting concept definitions and classifications.
5. Managing communication with end-users of statistical outputs and gathering of user feedback.

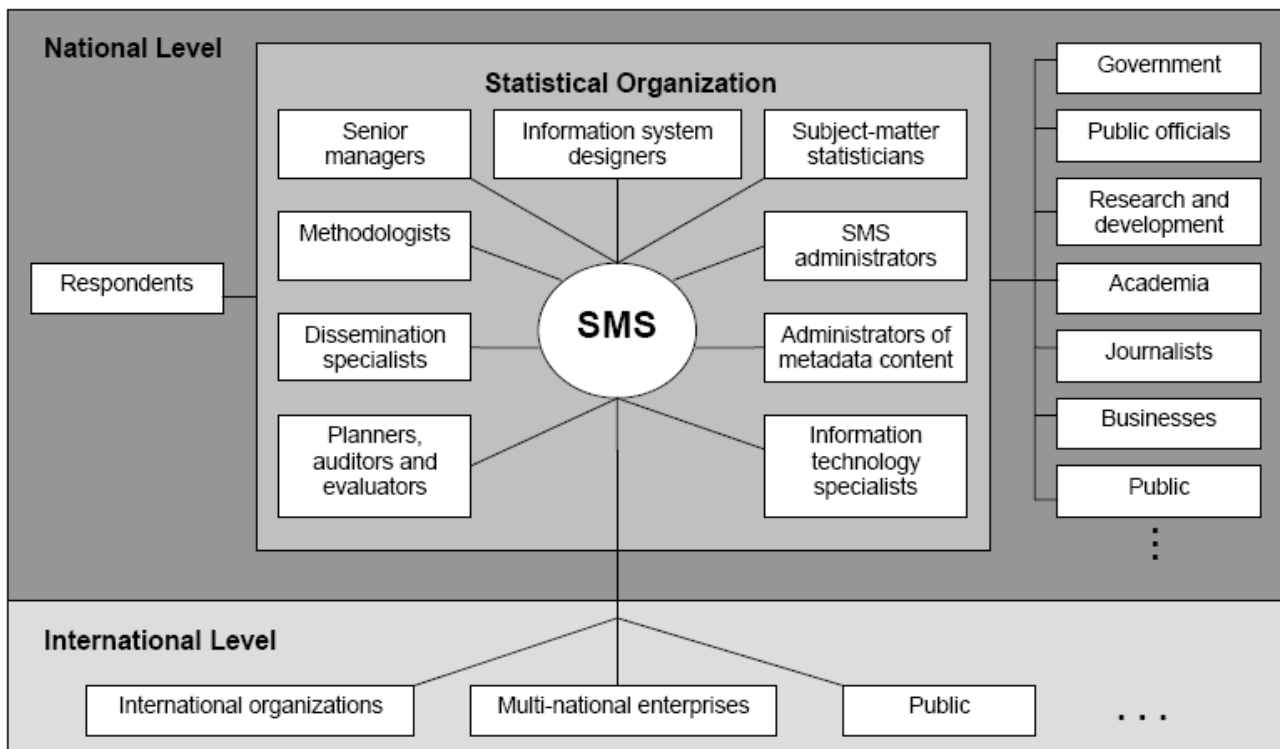


6. Improving the quality of statistical data and transparency of methodologies. Ensuring and evaluating the quality of statistical data is one of the most important activities. To this end, national and international statistical organizations have adopted a set of criteria (relevance and completeness, comparability and coherence of statistical concepts, accuracy of statistical estimations, timeliness and punctuality of delivered statistical information, its accessibility and clarity). The SMS should offer a relevant set of metadata for all of these criteria.
7. Managing statistical data sources and cooperation with respondents.
8. Improving discovery and exchange of data between the statistical organization and its users.
9. Improving integration of statistical information systems with other national information systems. Growing demands to use administrative data for statistical purposes require better integration and sharing of metadata between statistical and administrative bodies, to ensure coherence and consistency of exchanged information.
10. Disseminating statistical information to end users. End users need reliable metadata for searching, navigation, and interpretation of data. Metadata should also be available to assist post-processing of statistical data.
11. Improving integration between national and international organizations. International organizations are increasingly requiring integration of their own metadata with metadata of national statistical organizations in order to make statistical information more comparable and compatible, and to monitor the use of agreed standards.
12. Developing a knowledge base on the processes of statistical information systems, to share knowledge among staff and to minimize the risks related to knowledge loss when staff leave or change functions.
13. Improving administration of statistical information systems, including administration of responsibilities, compliance with legislation, performance and user satisfaction.
14. Facilitating the evaluation of costs and revenues for the statistical organization.
15. Unifying statistical terminology as a vehicle for better communication and understanding between managers, designers, subject-matter statisticians, methodologists, respondents and users of statistical information systems.

### 3. Users of the statistical metadata system

A primary challenge for the SMS is to cope with the requirements of diverse groups of metadata users. The use of evolving information and communication technologies has resulted in more users of statistics and a diversification of needs. Efforts should be made to understand who the users are, as their requirements for data and metadata may vary substantially.

Potential groups of statistical metadata users are illustrated in *Figure 1*. Each group is described below and the benefits an SMS will provide them are outlined in Section 4.



**Figure 1: Metadata users**

#### Users within the statistical organization

This group of metadata users encompasses the many professions involved in the phases of preparation, production and dissemination of official statistics and the functioning of the statistical information system. These include the following metadata users:

- Senior managers
- Information system designers
- Subject-matter statisticians
- Methodologists
- Administrators of metadata content
- SMS administrators
- Information technology specialists
- Dissemination specialists
- Planners, auditors and evaluators

## **Respondents**

Respondents are individuals or organizations that supply statistical data to the statistical information system. Responding to data requests often involves interpreting definitions and other metadata, in order to provide the required information. In this sense, respondents are important users, but where they provide additional information about the data they supply, they can also be considered as suppliers of metadata.

## **End users at the national level**

This group includes: government institutions, political decision makers, researchers, public officials, archivists, academics, librarians, journalists, businesses and the general public. As the number of end users grows, so does the range of their needs.

In the past, data dissemination methods assumed a certain level of knowledge about economic and statistical concepts. Frequently, users' knowledge for a given set of statistics was comparable to the knowledge of the subject-matter statisticians involved in producing it. This is no longer the case. The audience for economic statistics, for example, can range from professional economists and policy makers, to interested members of the general public, to students working on school assignments. An understanding of economic and statistical concepts can no longer be taken for granted. Improving statistical literacy is part of the work program of many statistical organizations.

## **International users**

International users of statistics, such as multi-national enterprises, international organizations, non-governmental organizations and others, are important users of statistical metadata. International standards, such as the Metadata Common Vocabulary, increase the usability of statistical information for these types of users. International standards also help to reduce the burden on national statistical systems when reporting to international and intergovernmental organizations.

## 4. Benefits for users

An effective SMS can provide the following benefits to all users of statistical metadata:

- Better quality statistical information;
- Improved interpretability of statistics;
- Improved quality of metadata;
- Better discovery, retrieval and exchange of data and metadata;
- Common terminology, names and descriptions for standard metadata elements to improve communication;
- Improved efficiency through central metadata repositories that are organized to facilitate reuse of existing data;
- Improved knowledge of metadata flows.

An effective SMS allows organizations to be flexible and responsive to rapidly evolving requirements for statistical information. Statistical organizations can interoperate more effectively with their respondents, with data consumers and with other agencies that form part of the larger statistical system. Well-maintained metadata allow the organization to operate in a more transparent and quality assured manner. The SMS supports more effective change management processes, reducing risks to business continuity and the barriers to business process improvement.

Statistical organizations and other stakeholders can benefit when metadata exist throughout the entire statistical production process, rather than as captive to a particular statistical processing system or infrastructure package.

### Users within a statistical organization

#### Senior managers

In addition to the general benefits described above, the SMS provides senior managers with a tool to facilitate the design, planning, decision-making and evaluation processes of statistical information systems, and enables other strategic initiatives to be designed and implemented much more effectively, efficiently and confidently. An SMS should provide the tools for answering questions like: to what extent do users actually use the statistical outputs? Are they satisfied with the quality of data and metadata? Are there complaints or unmet demands from users or respondents?

An SMS should provide answers to these questions, facilitating administrative management of a statistical system. Last but not least, senior managers of statistical information systems will be able to verify the costs and benefits of individual statistical activities.

To achieve these benefits, metadata about the following will be needed:

- End-user needs and stakeholder requirements on a national and international level;
- Available statistical services;
- External information systems related to statistical information systems;
- Suppliers and sources of data in statistical information systems;
- Statistical production process;
- Statistical publications, release calendar, copyright and other dissemination issues;
- Responsibilities inside the statistical organization;
- Costs and revenues.

### Information system designers

These people are responsible for the design, implementation, maintenance and evaluation of statistical systems. When designing new systems, they need access to metadata from existing systems, either within or outside the organization, to inform the design, development and implementation of a new system. For existing systems under their responsibility, they need feedback about performance, costs, usage, and user satisfaction.

The following information is required when designing and developing an SMS:

- How similar systems have been designed in the past;
- What observation data are already available;
- How can these data be obtained;
- What methods, tools and software components are available and how can they be used.

For maintenance and evaluation of an existing SMS, the following information will be needed:

- Detailed, up-to-date documentation of the system;
- Feedback, both formal and informal, concerning production and usage of the system;
- Experiences from similar systems;
- Knowledge about methods, tools and software components;
- Special evaluation studies performed on an ad hoc basis.

### Subject-matter statisticians

The subject-matter statistician is the expert in a particular field of statistics within a statistical organization. They have the crucial role of understanding the users' information requirements, in the context of the users' policy and program environment, as well as the capabilities of their statistical office, i.e. what they can do to provide the required information. Subject-matter staff work with other specialists to design and construct an appropriate collection mechanism and produce statistics. However, the statistician then has the role of communicating the information to users through the creation of statistical products and the provision of associated metadata to assist users in understanding the results. Evaluation is also an important responsibility for the subject-matter specialist.

Given these roles, the SMS is a knowledge management system for the subject-matter statistician. In this information system they would want to be able to create, update, search, browse and retrieve many different types of metadata entities that would cover many aspects, such as:

- user (customer) requirements;
- standard concepts, data elements and classifications;
- operational information and quality metrics about the operation of their survey system;
- documentation about statistical techniques (methodology) applied to their survey;
- products created from the statistical data.

The benefits of an SMS to the subject-matter statistician include:

- access to a consistent store of standard classifications, data elements, process engines that can be used in new statistical process development with the knowledge that using these elements will assist greatly in ensuring statistical integration;
- tools and links that enable the subject-matter statistician to create statistical products for the organization with a common 'look and feel';
- a record of statistical collections, including all previous cycles, and a reference point to find information on related collections. This is an invaluable resource for new employees coming

into a statistical field and for statisticians in other fields who might be researching a new collection - there may be elements in another statistical activity that can be reused;

- standard processes, such as the registration of new data elements, which would provide a common method for the creation and use of metadata;
- registration of who is making use of which metadata, and noting of any issues encountered with any of the existing metadata content. This facilitates effective information sharing between subject matter areas. It also facilitates change management for metadata which recognizes and supports the stakes held by each subject matter area.

## **Methodologists**

An SMS creates a framework for design and implementation of statistical tasks and surveys to meet obligations in the production of official statistics and needs of end users. The SMS provides tools for safeguarding the integration of statistical information systems at national and international level. It is an indispensable tool for the maintenance, use and further development of statistical classifications and nomenclatures; statistical registers; statistical standards; knowledge about statistical methods and relevant research methods.

Methodologists require metadata relating to the following:

- Content of available statistical data (microdata, macrodata) and associated data concepts;
- Quality of statistical data (relevance, accuracy, timeliness, punctuality, accessibility, clarity, coherence and comparability);
- Existing statistical tasks and surveys (questionnaires, other sources, etc);
- End users and their feedback;
- Requests of international organizations and related standards;
- Data sources and their links;
- Respondents' information systems;
- Administrative data;
- Information systems and their output databases (portals);
- Statistical registers (population, farms etc);
- Statistical classifications, nomenclatures and related international standards;
- Statistical population, statistical units, measurement units, time series;
- Statistical methods and relevant research projects.

## **Administrators of metadata content**

The SMS should ensure smooth and systematic update and maintenance of statistical metadata. Maintenance of metadata content will be performed by subject-matter specialists, methodologists and standards/metadata specialists responsible for metadata content. Metadata should be updated once and in one place. This will help avoid inconsistencies and unnecessary redundancies. Updates to all the dimensions of the corporate metadata repository should be automated.

The administrator will need a user-friendly interface, avoiding any special technical skill. The system should assist in identifying stakeholders which will be impacted by any administrative action, and in assessing the impact of that action on their use of the metadata. This assists in change notification, stakeholder consultation and risk/impact remediation. The administrator will need the following metadata:

- Information related to the content of and links between statistical metadata;
- Information about organization of metadata in the corporate metadata repository;
- Metadata allowing discovery and retrieval;
- Updating methods and procedures.

## **SMS administrators**

These people are responsible for the technical maintenance of the SMS. They should cooperate with designers, evaluators and content administrators in solving technological issues and for the further development of the SMS. The technical administrator will use, oversee and maintain the following metadata:

- Technical metadata related to the SMS, and to the links with production systems;
- Information and knowledge about technological aspects of statistical production;
- Information about technical links to other information systems;
- Information about tools and software used by content administrator.

## **Information technology specialists**

People operating and monitoring the statistical production process are important metadata users. Ideally, the SMS supports tuning of statistical business processes. For example, the statistical impact of a particular process (or choosing a particular threshold for that process, such as significance editing) can be assessed so that the practical “value added” of the process can be weighed against its costs in terms of resources and time. This is not so much tuning how the process works with metadata, but using metadata to tune the process.

Metadata driven statistical production creates favourable conditions for standardization and thus efficiency of statistical production systems. Metadata on the content of statistical data and associated concepts, including all other delimiting metadata (statistical classifications, statistical units, measurement unit, time series, statistical population etc), are a key condition for the whole throughput of production phases (data collection, processing, analysis and dissemination). Technical metadata on the organization of the corporate metadata repository, and links to the production systems, belong to the metadata set needed for fulfilling functions of data processing. The Generic Statistical Business Process Model ([www.unece.org/stats/gsbpm](http://www.unece.org/stats/gsbpm)) can be used as an organizing framework for metadata, as well as a means of benchmarking business processes.

Ideally, statistical production processes will generate metadata about their own performance, giving producers feedback about functioning and efficiency of metadata driven production. In this respect, producers should cooperate with SMS designers, subject-matter specialists and methodologists, content and technical administrators on the design, implementation, evaluation, and further development of the SMS.

## **Users outside the statistical organization**

### **Benefits for respondents and data suppliers**

Respondents are important partners of any statistical information system. Statistical data suppliers are often also the users of statistical data. Their role is becoming more important with the growing number of systems and online communication possibilities. Bearing in mind the possibility of electronic data reporting from respondents' information systems to the statistical information system and the possibility of online access of respondents to the statistical information system, it is evident that the needs of respondents will change. The SMS will play a key role in those tasks.

There is a growing need to harmonize methodological definitions of data and related metadata from respondents and statistical information systems. Attention should be drawn to the implementation and use of relevant technological metadata standards. SDMX (Statistical Data and Metadata eXchange) has been developed specifically for exchange of statistical data and metadata. The SDMX standards and guidelines aim at establishing a set of commonly recognised rules, adhered to by all players. This makes it possible not only to have easy access to statistical data, but also metadata, making the data more meaningful and usable. The standards will allow statistical organizations to fulfil their responsibilities towards users and partners, including international organizations, in a more efficient way, among other things by using their online databases to give access as soon as the data are released.

Respondents and data suppliers will require the following information:

- Metadata related to the content (definitions, terminology) of statistical data in the input stage of the statistical production;
- Security and confidentiality of microdata;
- Feedback from statistical outputs;
- Information about the content of statistical warehouses;
- Knowledge about comparability of statistical and respondents data/systems;
- Technical parameters for search and retrieval of metadata in the common metadata repository, and links to statistical warehouses;
- Knowledge about potential interface between statistical information systems and respondents' information systems;
- Relevant technological standards for metadata and data supply;
- Information about software and other tools supporting supply of data and metadata;
- Information about strategies for further SMS development;
- Training in use of the SMS.

### **Benefits for end users on the national level**

Understanding and classifying different communities of end users could help in determining user requirements. The SMS will help users to better discover, understand, interpret and interrogate the data they need. The proliferation of statistical information has raised the issue of consistency and comparability of data. Comparability of data is desirable, but not always possible. It is important to know what the differences are and the reasons for them, with explanations according to differing levels of user understanding of statistical concepts. The SMS will also assist to convey the credibility of statistical data and to recognize intellectual property rights.

It is important to monitor user feedback and to embrace the need for metadata in both directions. The SMS will offer the possibility to understand how users search and the terms that they use. The SMS will also support the management of access to microdata. The fact that users are increasingly requesting access to microdata, calls for tools that allow concerns about confidentiality protection to be overcome.



With the spreading use of the Internet it is important to provide users with the appropriate information about the data available from statistical websites. However, there is a potential to flood users with too much metadata. Appropriate communication of metadata should be based on principles of 'cognitive psychology', recognizing the important role that presentation plays in metadata consumption.

An increase in the possibilities for syndication and reuse of data by external websites, such as online communities, web services and 'mashups', means that metadata need to be more closely, but flexibly coupled with data in a way that both web services and people can use. The websites of statistical organizations may not be the main source for data, with users going to an increasing number of secondary providers of online statistical information.

This heterogeneity, together with more visible methodological differences and inconsistencies of statistics disseminated via the Internet, poses difficulties for the users. Clearly, there is a need for harmonization of metadata accompanying statistical information on the Internet. International standards should play an important role in this respect.

The following metadata are vital for end users of statistical metadata and data:

- Availability of statistical outputs;
- Metadata related to the statistical outputs (metadata and data concepts and definitions, classifications, aggregations, statistical and evaluation methods, terminology, history, etc);
- Metadata about quality (e.g. explanatory notes);
- Access to microdata;
- Time series;
- Updating procedures;
- Statistical revisions;
- Responsibility for individual statistical outputs;
- Links to other information systems both national and international;
- Confidentiality;
- Planned changes in statistical outputs;
- Content related standards, both national and international;
- Outcomes from statistical analysis on users feedback;
- Rules for searching, accessing and downloading statistical metadata and data from output databases;
- Technological standards relevant for extraction and transfer of data and metadata;
- Information about software and other tools supporting search, retrieval and downloading of metadata and data;
- Users training possibilities;
- Metadata based services such as classification coders and metadata mappings that other producers and users of statistics can apply.

### **Benefits for international users**

There are more and more demands by international users for greater consistency when interacting with statistical organizations. In the case of international organizations, the metadata and data requirements (particularly concerning collection and exchange) have to be coordinated not to overburden countries with duplicate requests. In order to fulfil this task, better integration of metadata at the national and international level is needed.

A lot of metadata are available on websites of international organizations. Links could be inserted from the metadata of international organizations to more detailed metadata on national websites. Coordination of access could be achieved through a single gateway for data and metadata. To this end, joint hubs based on SDMX standards are at present under intensive development.

The needs of international users increasingly impact the architecture of the SMS of national statistical organizations. International collaboration and alignment should be driven as much by national statistical organizations – and their national interests - as by international organizations. Processing and storage power allows formats that are globally rather than locally optimised (e.g. XML) to be viable, opening up practical application of standards. This is coupled with movements and standard, open, toolsets that promote “open source” and similar collaborative developments. There are unprecedented opportunities to collaborate. At the same time there are unprecedented rates of change in technologies, end user expectations, information needs and data sources, which require national statistical offices to collaborate in order to continue meeting user requirements.

Metadata needed by international users are similar to those needed by end users on national level (see above). Furthermore, the following information would be required:

- Complying with international standards (coherence, comparability, explanatory notes);
- Standards used for electronic metadata and data transfer;
- Information about other international and national users;
- Indication of needs for revision and/or standardization of statistical data and metadata concepts.

## 5. SMS vision, strategy and implementation

The focus of this section is on the preparation of a corporate SMS vision, related planning and on the major characteristics of a metadata management framework and management strategy.

### SMS vision

The vision should clearly state the goals or aims of the SMS. It should apply across the entire statistical information system and be realistic and within the capabilities of the statistical organization. It should also include a statement about scope: what is included in the SMS and what is not.

The SMS vision should support and clearly demonstrate its alignment with, and relevance to, the overall corporate vision and priorities of the organization. The SMS will be an essential enabler to achieve broader business, statistical and technical outcomes within the organization. In other words, it serves as a means to many ends, rather than being an end in its own right.

### Vision goals

An important prerequisite for successful design, implementation and functioning of the SMS is the development of a corporate vision of the SMS in the statistical organization. The functions of the SMS, centred on metadata and data users, should be oriented towards the diverse processes and activities of the statistical information system. Operational units within a statistical organization, respondents, end users and other stakeholders should all have an input to the SMS vision, and, where practical, should be involved in the preparation, implementation, use and evaluation of the components of the SMS.

The vision should be an integral part of the strategic direction of the statistical organization, so should be developed with the direct involvement of senior management. It is an important task for the SMS managers to ensure that not only the development of the vision but also the SMS design, implementation and further development will be actively supported by senior managers. For this purpose a suitable SMS management and reporting structure should be established. Feedback and evaluation, supported by metadata accumulated in statistical production cycles, should be an integral part of the SMS design.

The vision should define major goals and functions of the SMS for the statistical organization and attribute priorities for their implementation. This includes clearly linking the goals of the SMS with strategic business drivers from within the organization and beyond.

The metadata requirements associated with each part of the statistical business process should be identified. All points of contact between the SMS and business processes, in terms of creation, update and use of metadata should be described. It is important that the high level picture can be discerned, along with the details. The Generic Statistical Business Process Model ([www.unece.org/stats/gsbpm](http://www.unece.org/stats/gsbpm)) provides a tool for identifying at least higher-level links. The points of contact and relationships between the metadata model and business processes should be pictured both as they currently exist and as is envisaged in the future. This will summarize for the organization what would change, both at the strategic and the day-to-day level, as a result of realizing the vision. This, in turn, identifies the benefits more clearly from a business perspective.

An important part of the vision should be an analysis of the existing statistical metadata objects and services, finishing with a clear specification of which existing metadata and services can be used in the corporate SMS, which should be updated and which should not be used at all.

It is advisable, that the SMS is not developed as a purely technical project. It is still quite often the case in statistical organizations that the subject-matter departments do not understand fully the requests formulated by information technology specialists. When developing the vision, it is essential to express clearly that the first priority in the SMS is to safeguard the content and logical integration of statistical data and metadata.

To make the SMS a success story, the vision and its implications should be based on what is really possible for the organization. Effective management of the statistical information system and the integration of information flows, on both the national and international levels, should remain one of the major goals of SMS.

The vision should also encompass cost estimates for the SMS project. Wherever possible, tangible benefits should be identified and weighed against these costs. It is also recommended that likely medium to longer term costs and risks of inaction are estimated (e.g. ongoing duplication of effort and information, ongoing complexity and inconsistency, ongoing cost and risk of inertia to evolution to keep pace with changing statistical requirements). These should be broadly compared with the costs and risks of taking action. It should be ensured that the cost to producers of metadata is justified by the benefits to the metadata users. Such proposals should be very pragmatic, reflecting ultimate needs and metadata priorities. Experience shows that the human and financial costs of SMS developments can be quite demanding.

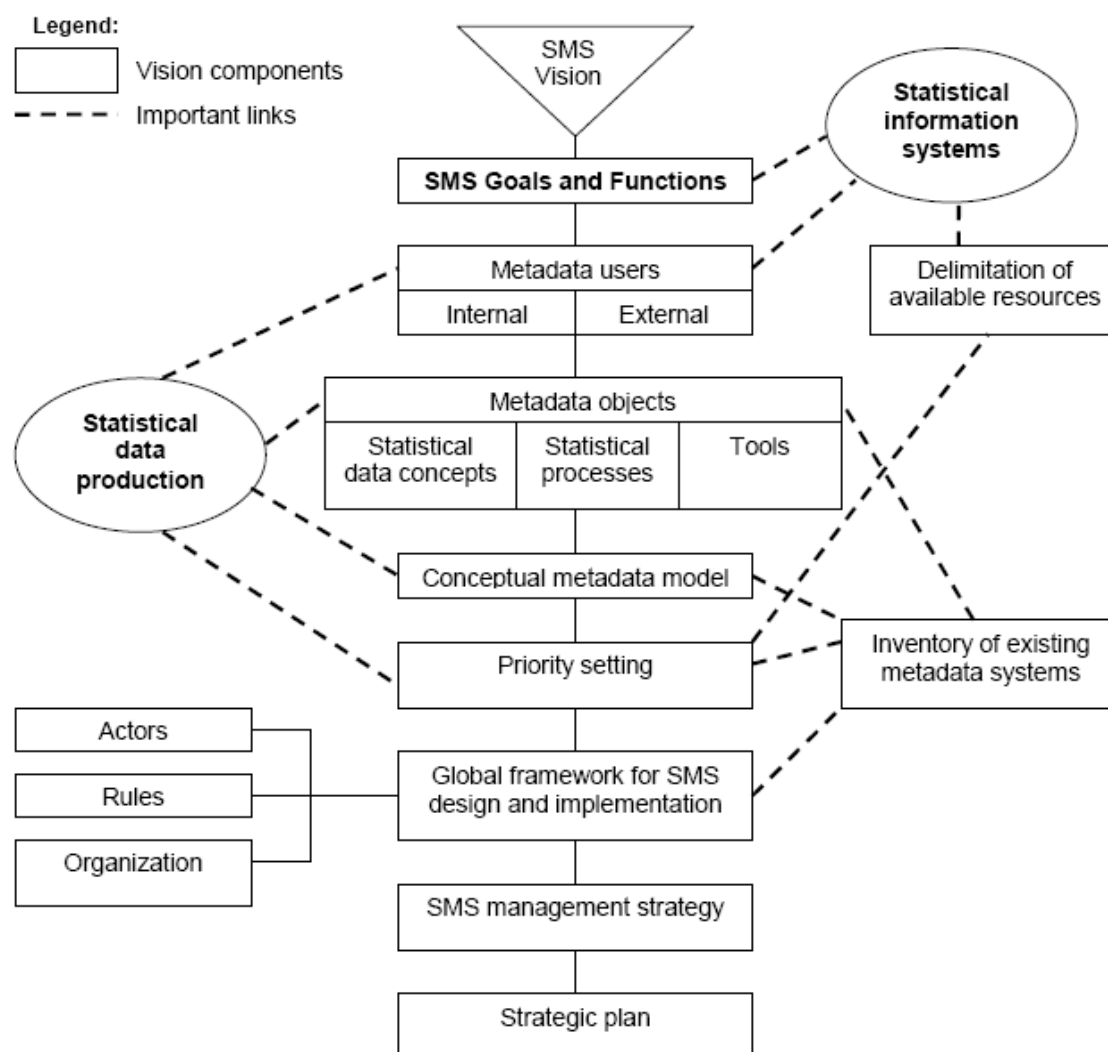
Experience also shows, that many statistical organizations implemented some functional blocks of metadata without having a complete SMS vision at the beginning of the process. This includes lack of a practical strategy for how different “functional blocks” delivered at different times, possibly by different teams, will knit together into a coherent SMS capability and how this capability will be harnessed in practice by all the different aspects of the statistical information system. It is especially true for the objects dealing with the description of statistical data. It can be observed, that the following blocks of metadata have been frequently implemented: statistical variables and values sets, statistical surveys, socio-economic classifications and nomenclatures, time series, statistical publications, statistical population, economic subjects, statistical units, aggregation and statistical evaluation methods, output tables and others.

Without having a coherent vision there is very often a lack of coordination among individual metadata blocks. This causes many inconsistencies, duplications and, last but not least, the low efficiency of metadata tools in terms of both costs and staff capacities needed. The end users could, because of lack of coordination, struggle with unnecessary diversity of users’ roles and related diversity (and possibly inconsistency) of terms and concepts being used in different existing metadata blocks. Such a situation certainly does not encourage cooperation of users with metadata implementation.

The vision should contain a metadata model complying with the SMS functions. Such a model should encompass metadata about data and processes behind them as well as metadata about other objects and processes of the statistical information system relevant to the SMS functions. Metadata needed for the management and administration of the statistical system such as metadata about costs and benefits, cost-effectiveness, satisfaction and complaints should also be a part of such a model. Metadata objects and links between them should be identified.

An agreed conceptual metadata model should be linked to the standard business processes that are the part of the statistical business process. This linkage is used to determine which metadata should be collected, and at which point in the process. Metadata models should take account of and use international standards where possible.

Figure 2 below provides an overview of the components of the SMS vision.



**Figure 2: Schematic view of the SMS vision and its components**

### Metadata objects and metadata resources

Metadata should be structured according to the objects they describe. There are three main categories of objects to be considered when preparing the vision:

- i. statistical data
- ii. statistical processes
- iii. tools enabling production and use of statistics

### Statistical data

Metadata related to statistical data are important tools supporting the production and final use of statistical information. These metadata include descriptions of statistical concepts, characteristics, variables, units, populations, classifications, registers, observation templates, statistical surveys, time series, observations, aggregations, methods, micro data, macro data, final outputs, statistical publications, statistical databases and archives. Information about respondents, end users, statistical websites and other metadata objects related to statistical data, also belong to this group.

## Statistical processes

Statistical processes can be divided into two major groups:

- i. those associated with statistical production (data collection, data storage, data evaluation, data dissemination); and
- ii. those associated with the statistical information system and statistical organization (planning and evaluation processes, supply processes, usage processes, total quality management and other management processes).

All processes are associated with important metadata, such as information on costs, performance measures, errors and error rates, diverse benchmarking indicators, etc. Process metadata are sometimes also referred to as paradata.

Processes are related to metadata in three ways:

- i. they are carriers of metadata, passing them on to subsequent processes;
- ii. they use metadata;
- iii. they produce metadata.

Figure 3 illustrates the links between the metadata objects described above.

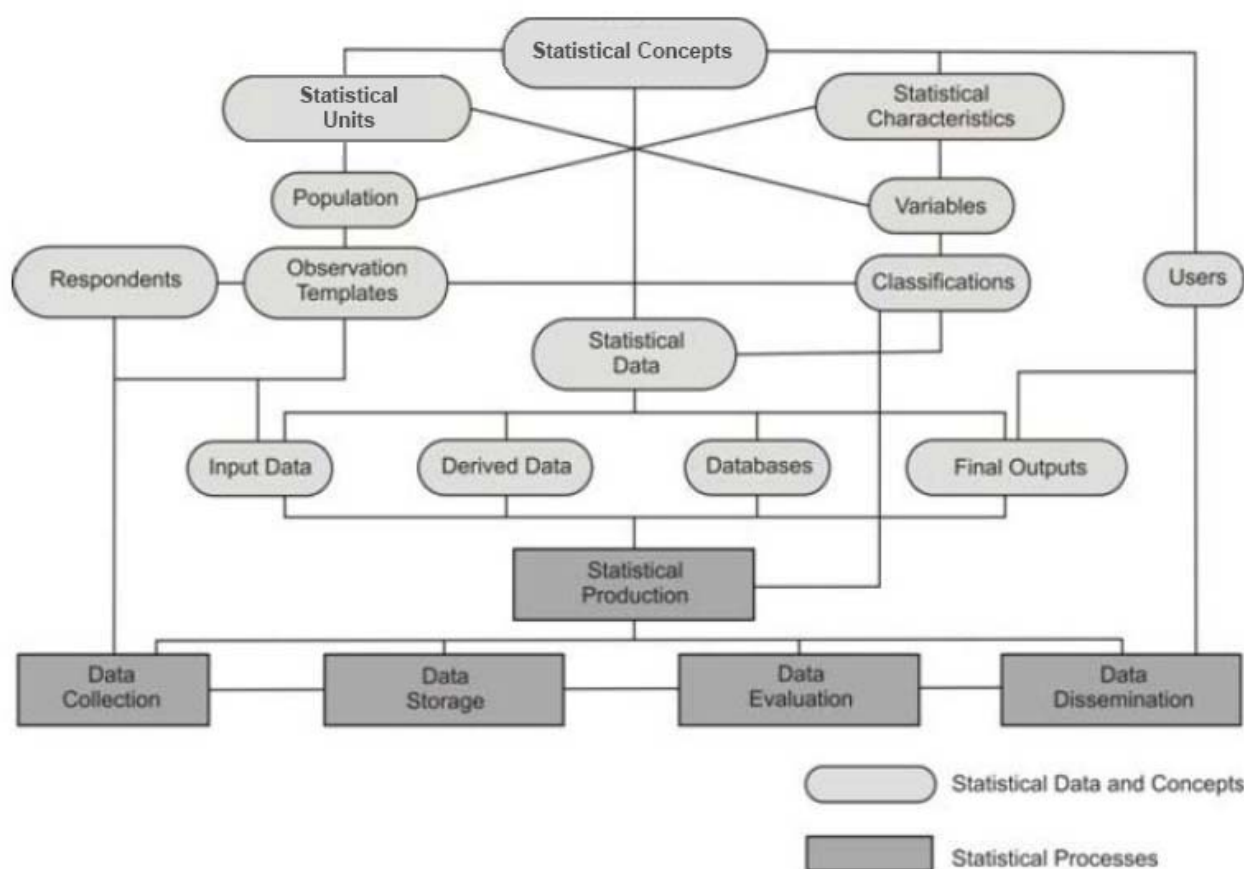


Figure 3: Example of metadata objects for statistical production processes

## Tools enabling the production and use of statistics

The SMS should provide tools to enable the production and use of statistics including: (i) search and retrieval tools; (ii) tools supporting statistical production, and; (iii) knowledge resources supporting the “intellectual processes” related to the statistical system, such as corporate management, planning and evaluation, research and development. These tools should be sharable by multiple processes. They need to be systematized and organized collectively in order to be easy to find and make use of, both by people and by automated processes within the statistical information system.

In this respect, the vision should promote the following:

- i. Development of common terminology for metadata elements across all processes in the statistical life-cycle;
- ii. Development of a common and consistent description of metadata elements allowing easy location, retrieval and exchange of data and metadata;
- iii. Development of standard interchange formats and tools allowing sharing of metadata and data between organizations;
- iv. Implementation of consolidated metadata repositories facilitating reuse of metadata;
- v. Introduction of a registration processes to promote the use of standard metadata elements and increase knowledge on metadata flows and statistical integration;
- vi. Improvement of metadata quality;
- vii. That the production process will be metadata driven.
- viii. Clear policies on how organizational and technical processes outside the SMS will change once it is agreed that the SMS is fit for purpose, and capabilities that allow these changes to take place. What constitutes “fit for purpose” (i.e. not necessarily perfect) should be agreed beforehand.

## SMS planning

### Preparation of the strategic plan

The aim of this subsection is to advise on the preparation of a corporate strategic plan for the development of an SMS. A strategic plan should be an integral part of the SMS vision, reflecting the goals and functions specified in this document. As a part of the vision, the senior management of the statistical organization should approve the strategic plan.

The development of a strategic plan needs to be a flexible and adaptive process, possibly with several iterations. The plan should give visibility, clarity and stability to the development efforts, but aspects are likely to change during its implementation, which may take several years. Certain parts may never be implemented; other parts may be implemented in a different way than originally assumed. Completely new components may appear as a result of new needs, new methodological and technical developments or changes of some other basic conditions for SMS development. Therefore, the plan should be regularly reviewed and revised.

Detailed plans should be developed and approved later on for the design and implementation phases of the SMS development. Such plans should reflect agreed priorities for the development of individual components of the SMS. Last but not least, specific plans should be prepared for the use and evaluation of the SMS components.

When preparing a strategic plan, the number of activities, sensitivity of their solution and their priorities for the statistical organization should be taken into consideration. Links between individual activities and importance of their contribution to the SMS strategic goals should be thoroughly analyzed. Conditions under which the goals could be met should be clearly specified.

A key part of the plan should be the establishment of an organizational framework and management strategy.

The strategic plan should be developed and approved by all actors involved in the design, implementation and maintenance of the SMS. It is therefore indispensable that such plan is prepared in close dialogue and cooperation with all actors involved in the process of the SMS development. The planning should be made explicit, so that the whole organization can discuss the strategies to be used and the choices to be made in the step-by-step development of the SMS. At the same time as the step-by-step development of the SMS is being planned, the subsequent uptake and use of the capabilities to be delivered should also be planned.

### **Recommendations for establishing the strategic plan**

The recommendations below are based on the experiences of several national and international statistical organizations:

- i. When preparing the plan, the organization should consider its current capabilities. Available human and financial resources, as well as organizational and technical feasibility, should be carefully analyzed in order to make the plan realistic.
- ii. Goals defined in the vision should be transformed into practical steps to which priorities are then assigned.
- iii. Practice shows that different organizations often have similar priorities. This is especially true for the development of databases on statistical classifications and nomenclatures, aggregated output databases, and metadata models for the websites. Some organizations give priority to metadata models for microdata. Discussions and sharing of experiences between statistical organizations are therefore strongly recommended.
- iv. Quality of data and metadata should be considered a high priority.
- v. External cooperation should be clearly defined; categorization and priority setting for external users should be specified. The plan should take the existing working plans of all external partners into consideration.
- vi. The plan should be prepared in such detail that all partners will be able to commit their participation.
- vii. External projects to establish data and metadata warehouses, both on the national and international level, should be considered for potential impact on the SMS.
- viii. External activities on data security and data confidentiality related to the SMS should be considered.
- ix. An integral part of the plan should be activities dealing with the development and implementation of international standards.
- x. The plan should also consider activities to promote the SMS and create an atmosphere of cooperation with all stakeholders. To this end, prototypes for demonstration of SMS functions could be useful.
- xi. Research activities on feasibility studies and analysis of user feedback should be also taken into the consideration when preparing an SMS plan.
- xii. Transfer of know-how and training for participants in the SMS business case should be incorporated in the plan.



## Management strategies for corporate SMS

A framework for a corporate metadata management strategy should be specified in the vision. The senior management should play a lead role in a corporate management model.

Responsibility for development of metadata policies and procedures and for providing training and advice to developers should be clearly assigned.

An important part of the SMS management strategy should be a systematic cooperation with major metadata stakeholders.

Implementation of the metadata management strategy should follow two broad approaches. They are:

- i. User orientation – focusing on information relevant to usage such as finding and accessing data, understanding their structure and meaning, assessing their quality and relevancy, and using them correctly. This focus is dissemination oriented; and
- ii. Producer orientation – metadata driven approach focusing on the needs of information systems and electronic processing.

There are two major dimensions to be taken into consideration when deciding on SMS management strategy: (i) the crosscutting nature of the SMS role and its functions in statistical organizations and, (ii) the requirement of corporate management during all phases of SMS development and use.

## SMS management across the whole statistical organization

The SMS is an integral part of a statistical organization's strategic direction. It implies that the SMS management strategy should be integrated into the management strategy of the organization.

Diverse organizational units and external bodies participate in the SMS business case. Managers, subject-matter statisticians, methodologists, information technology experts, researchers, respondents and end users are all SMS partners. Their functions, needs and obligations differ according to whether they participate in the SMS as metadata users, metadata suppliers, designers, developers, producers, administrators and/or evaluators. Clearly, the SMS does have a cross-cutting nature. The management strategy for the SMS business case should correspond to those needs.

It is strongly recommended that the top management of the organization is directly involved in the SMS and its management, but experience shows that this can be hard to achieve.

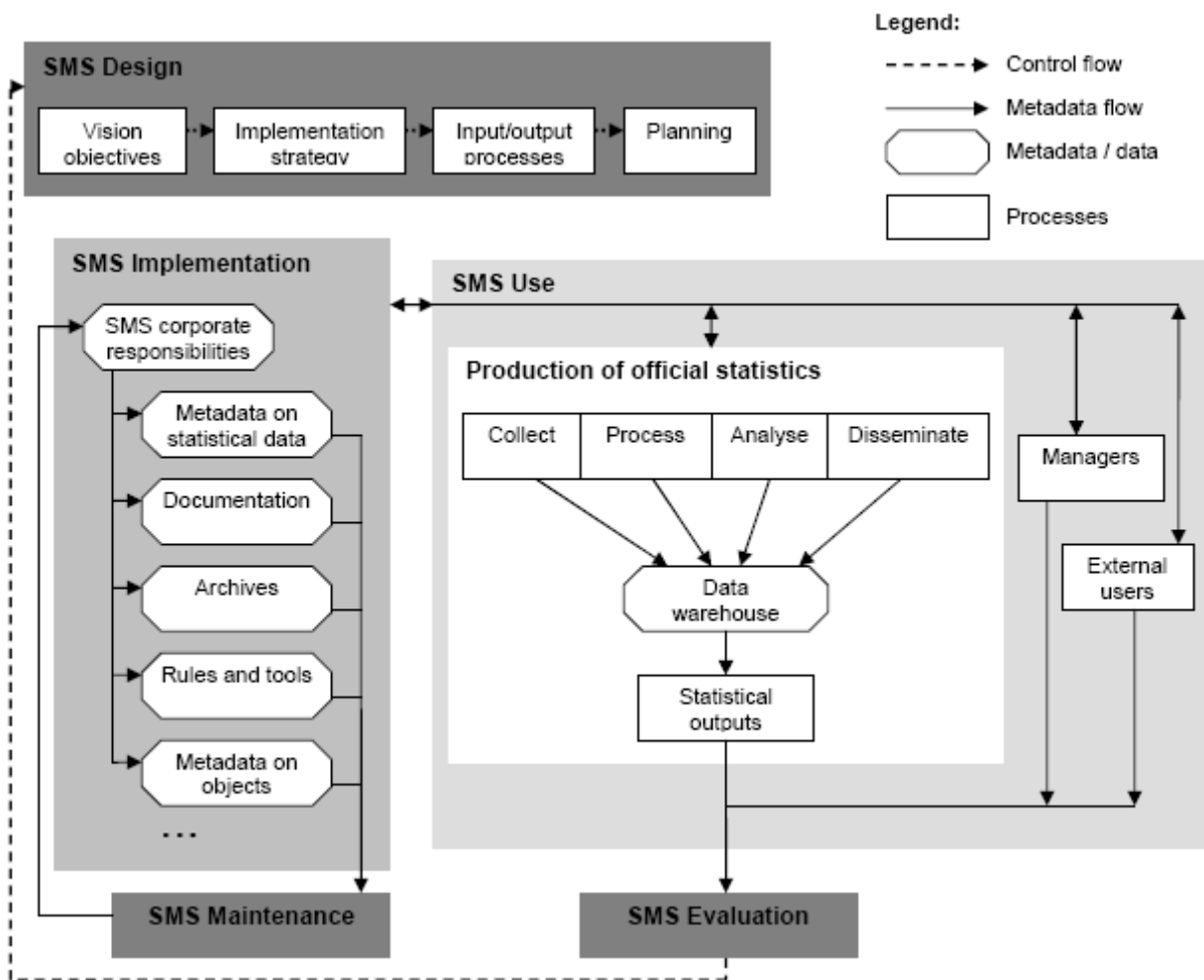
Some recommendations for the SMS management strategy across the whole organization:

- i. Metadata management is a part of every project and should be considered alongside resource allocation and accountabilities, in the same way as business processes and data flows are considered.
- ii. The SMS management strategy should be specified in close alliance with the existing managerial structure of the organization. With the lead role of the senior management in the SMS management model, clear links should also be defined in the middle management level and in the experts' level (methodologists, subject-matter statisticians, information technology experts).
- iii. Roles and responsibilities of all partners should be clearly defined, understood and followed. Where possible, automated workflows can be used to enforce agreed roles and responsibilities.

- iv. An SMS management board should be established. This board will take an ultimate, corporate view on all decisions dealing with the SMS development.
- v. A multidisciplinary team should be the major organizational form for the development of the SMS project. The “ideal” SMS Team(s) will include: statistical methodologists; subject-matter statisticians, dissemination specialists, end users, standards’ experts, researchers, and information technology specialists in data modelling, business process design, architecture and applications development.
- vi. Implementation of the SMS management strategy may highlight some needs for changes in the job description of some experts (namely methodologists and subject-matter statisticians) as well as in the organization of statistical work. This will often be the case in statistical organizations where a corporate SMS did not previously exist. Many critical issues could appear. Such issues should be foreseen and reflected in the SMS vision and strategic plans as far as possible.

**Corporate management of the SMS development life cycle**

Figure 4 presents a model for managing the phases of an SMS development life cycle, being design, implementation, maintenance, use and evaluation. The governance of metadata management and the monitoring of outcomes should be made clear in the SMS management strategy.



**Figure 4: Model for integrated management of SMS**

The most important management activities in each phase of the SMS development life-cycle are outlined below.

### **Management of SMS design**

The role of the design phase is to develop the SMS vision and global architecture, and to establish a management and implementation strategy for the project. The most important functions, tasks and activities to be considered by management are as follows:

- i. Development of the SMS vision.
- ii. A global plan for SMS development should be established and approved by all participants.
- iii. To ensure the efficiency and value of metadata-related work, an SMS Global Architecture should be developed. This should encompass all processes that will work with metadata. An inventory of all such processes and existing metadata tools should be prepared. This inventory should be developed in close cooperation with major stakeholders. The order of priorities should then be decided.
- iv. Specification of common components: Analyzing the inventory should reveal common metadata components. The possibility for these components to communicate with different parts of the SMS via a single interface should be explored. This should be considered a high priority.
- v. The impact of the corporate SMS on existing statistical production systems should be clear from the SMS vision. Any necessary re-engineering of processes should be considered and planned at this stage.
- vi. The metadata requirements associated with standard business processes are articulated, i.e. all the points of contact between the SMS metadata model and business processes, in terms of creation, update, and use activities, are described. For example, metadata associated with understanding user needs, frameworks and standards is acquired and used to inform later phases. To the greatest extent possible, the necessary input and output metadata should be captured in the collection strategy stage so that we know well in advance that the desired outputs are obtainable and fit for purpose.
- vii. Major partners in the design phase are the users (both, inside and outside the organization), methodologists, subject-matter statisticians and information technology experts.
- viii. Feedback and evaluation is an integral part of the design process and is supported by metadata, accumulated in all phases of the SMS development life cycle.
- ix. During the transition period, some “legacy” metadata structures and processes may need to be supported by the SMS to meet requirements of legacy components of the statistical information system. The SMS architecture needs to clearly and consistently differentiate such legacy aspects from the aspects designed for the future to ensure that the latter are always applied in preference to the former.
- x. Financial requirements for implementation phase should be detailed.

### **Management of SMS implementation**

The role of this phase is to implement the SMS so that it is ready for use. The implementation of all SMS subprojects can be a long process. Depending on dependencies and priorities, some subprojects may be implemented in parallel and others sequentially.

The following major activities should be considered when preparing a strategy for this phase of development:

- i. An agreed set of definitions and terminology should be developed. Consideration of national and international terminology standards, such as the Metadata Common Vocabulary published by the Statistical Data and Metadata eXchange (SDMX) initiative, is of high importance.
- ii. Detailed and coordinated plans for all stages of SMS implementation should be prepared and approved by all partners at the beginning of the implementation phase. The basic framework of the SMS plan is defined in the vision.
- iii. Existing processes using statistical metadata should have been reengineered. If complete reengineering cannot be achieved in advance, then firm plans and project management for achieving it in a timely manner must be in place.
- iv. Outsourcing possibilities for the SMS implementation should be considered.
- v. It is recommended to implement an SMS as a technically coherent project. This will allow standard links between metadata objects and processes, standard metadata tools for searching, retrieval, exporting and downloading metadata and harmonization of technical administration. Standard operations for administration of diverse metadata can then be developed.
- vi. A crucial task in the implementation phase is to set up a corporate metadata repository. This is the physical implementation of the metadata model defined in the vision. The concept of the corporate metadata repository should be developed, although there could be a number of physical repositories. To develop an appropriate architecture is a demanding task and there is no blueprint for such an exercise. However, good practices exist that may be a useful guide.
- vii. Physical loading of metadata into the corporate metadata repository should be done by metadata owners. This is a resource-consuming task and the impact on subject-matter staff should be recognised. For many, capturing metadata is a tedious extra task that brings them no perceived benefit. A prerequisite of the 'system' therefore, is that as many metadata as possible are captured automatically, as a result of a computer process or as a result of a required business process undertaken by a person. Thorough management and planning of those activities is imperative. Management support and highlighting benefits to users (refer to Section 4) can help motivate them to provide the necessary metadata.
- viii. Regular monitoring of progress in implementation from the view of completeness and cost effectiveness is necessary.
- ix. Tools and processes specified in the vision should be developed and tested by all target user groups. User manuals and documentation should be developed. Testing should be conducted before making the SMS available. Training for all metadata users should be organized.

### **Management of SMS use**

The role of this phase is to ensure efficient use of metadata and metadata tools by all users specified in the vision. Effort should be made by management to monitor and coordinate activities and processes dealing with metadata usage by the broad range of users (refer to Sections 3 and 4). The metadata strategy in this phase should encompass the following functions:

- i. Prepare, maintain and coordinate detailed plans of metadata use by all users to ensure the required metadata quality within the required deadlines. The coordination of plans developed for individual users is a major goal of the management.
- ii. The units responsible for statistical production should be accountable for the preparation and maintenance of plans related to the activities dealing with the production process. In this case, the SMS management should ensure that all activities dealing with the use of statistical metadata and metadata tools are planned and defined.

- iii. Oversee the availability of metadata and metadata tools and ensure the links between metadata maintenance and use.
- iv. Inform metadata users about all changes to metadata contents.
- v. Arrange for ongoing feedback from users about metadata quality and the availability and efficiency of metadata tools. Feedback mechanisms could be integrated into regular activities of metadata use. Specially organized surveys on user satisfaction are useful, but should be complemented by these ongoing mechanisms.
- vi. SMS management (in close cooperation with the SMS technical administrator) should be aware of the software and technological environment related to the use of metadata and metadata tools. As mentioned previously, metadata and metadata tools should be platform independent. However, it could be useful to maintain information about changes in the users' software environment (e.g. updates to web browsers, etc).
- vii. Statistical websites are an integral part of SMS implementation and the use of metadata. Furthermore, they are a regular part of the dissemination strategy of a statistical organization. The structure and quality of metadata presented on the website contribute to the satisfaction of metadata users. The need for statistical metadata on websites varies according to the needs of the individual users groups. The use of statistical metadata on websites should be monitored in order to keep track of users' satisfaction and evolution in their needs.

### **Management of SMS maintenance**

The role of the SMS maintenance phase is to ensure that all metadata stored in the corporate metadata repository are up-to-date for ongoing use. To keep metadata up-to-date is of primary importance for all metadata users.

The following recommendations should be taken into consideration when preparing a strategy for the management of this phase:

- i. The major functions to be considered are those relating to the administration of metadata content.
- ii. Planning is an important instrument for managing the maintenance phase. Everyone participating in maintenance processes should approve a detailed plan of maintenance activities, which meets required timelines. Such a plan is an indispensable instrument for management for smooth and coherent monitoring of the phase of metadata maintenance.
- iii. Ensure timeliness and coherence of maintenance activities.
- iv. The concept of registration of metadata objects, ownership of metadata, what is the 'standard' for a particular classification or data item, what are the permitted variations from the 'standard' etc should be all clearly defined, agreed and used.
- v. SMS management should oversee the definition and maintenance of all metadata stored in corporate metadata repository, although other units in the statistical organization will also contribute to its ongoing enhancement.
- vi. SMS management is responsible for definition of policies, procedures and protocols for the maintenance of the corporate metadata repository. A 'registration authority' manages all metadata entities in corporate metadata repository. The major partners for the SMS management are the "custodians" of metadata. The custodians are, according to the concepts of registration of metadata object specified in the vision, authorized to keep the metadata they are responsible for up-to-date.
- vii. Rules and guidelines should be developed for the maintenance of each metadata entity in the corporate metadata repository, identifying the responsible metadata owner. It is recommended

that these rules and guidelines are approved by senior management and become official documents of the statistical organization.

- viii. Preparation of rules and guidelines requires joint work with metadata owners. Methodologists are also important partners in this process.
- ix. Training of metadata owners in the rules and guidelines prepared for maintenance activities will be required.
- x. SMS management should ensure maintenance of metadata history and updating of links between metadata in the corporate metadata repository.
- xi. All maintenance functions performed by metadata administrators and metadata owners should use a coherent/standard set of metadata tools. Such tools should be available especially for the following maintenance functions: search and retrieval, inserting and deleting of metadata objects and related parameters, changes and corrections, presentations and exports, metadata editing and consistency controls, checking and updating of metadata links, maintenance of metadata history.

### **Management of SMS evaluation**

The goal of the evaluation phase is to determine the efficiency of existing SMS functions and make proposals for improvement or further development of the SMS. There are clear links to the knowledge and experiences accumulated in the earlier phase of the SMS development life cycle, namely in the SMS use phase. By preparing proposals for further SMS development, the SMS evaluation phase makes a loop between the use and design phase of the SMS.

The management strategy of the SMS evaluation phase should include the following procedures and tasks:

- i. Specify the major targets of SMS evaluation and prepare a plan of evaluation activities and procedures based on these. It should be clear which functions and aspects of the SMS are to be evaluated.
- ii. Evaluation of user satisfaction should be an ongoing part of the SMS development life cycle. The most important object of evaluation is ultimately the external user of data and metadata produced by the statistical organization. For these users, it may not be appropriate to measure satisfaction with the SMS in isolation, as they are likely to see data and metadata as a single package. Incorporating questions on metadata and the external interfaces of the SMS alongside more general questions on satisfaction with data outputs may be more effective, and increases the potential for additional insights into user needs and satisfaction through cross-tabulation of results for different questions. It should be ensured however, that the satisfaction of other user groups, particularly internal users would also be evaluated.
- iii. Other important aspects for evaluation are cost efficiency, implementation of standards, organization of work, maintenance procedures and technological implementation.
- iv. In principle, there could be three major forms of evaluation: (i) regular long-term evaluations (e.g. at 3 year intervals) that examine overall effectiveness of SMS functionality; (ii) regular short-term evaluations (e.g. annually) that primarily assess user satisfaction; and (iii) ad hoc evaluations as deemed necessary.
- v. Benchmarks should be established for all defined targets. Benchmarking parameters and evaluation methods should be specified and agreed. For some cases an efficient benchmarking method is to compare experiences and plans with those of a similar organization. International cooperation can be highly efficient in this respect.

- vi. Appoint evaluators for planned evaluation activities. The team of evaluators should include both staff from the statistical organization and metadata users. For evaluation of the project's efficiency and the overall technological solution, it may be useful to hire external evaluators to provide an independent view.
- vii. Document information on the user feedback collected in the phase of the SMS use.
- viii. Organize specific surveys on user satisfaction.
- ix. Report to the senior management of the statistical organization on the evaluation outcomes, including prioritized recommendations. Based on the conclusions made by senior management, organize improvement of and/or further development of the SMS.

Evaluations should focus both on the effectiveness of the SMS as an output (and an element of statistical infrastructure), and on the effectiveness of the broader business outcomes being achieved. The latter should be compared with the original vision, but should also take into account any changes to business directions and priorities which have occurred in the meantime. As well as changes to the SMS itself, changes may need to be considered to implementation objectives, planning and governance.

## 6. Core principles for metadata management

This section focuses on the management of statistical metadata in the SMS framework. It presents the principles to be taken into the consideration when preparing the SMS vision, global architecture and when implementing the SMS.

The principles can be presented in the following groups:

- |   |  |
|---|--|
| <b>Metadata handling</b>                              | <ul style="list-style-type: none"> <li>i. <b>Statistical business process model:</b> Manage metadata with a focus on the overall statistical business process model (<a href="http://www.unece.org/stats/gsbpm">www.unece.org/stats/gsbpm</a>).</li> <li>ii. <b>Active not passive:</b> Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date.</li> <li>iii. <b>Reuse:</b> Reuse metadata where possible for statistical integration as well as efficiency reasons</li> <li>iv. <b>Versions:</b> Preserve history (old versions) of metadata.</li> </ul>  |
| <b>Metadata Authority</b>                             | <ul style="list-style-type: none"> <li>i. <b>Registration:</b> Ensure the registration process (workflow) associated with each metadata element is well documented so there is clear identification of ownership, approval status, date of operation, etc.</li> <li>ii. <b>Single source:</b> Ensure that a single, authoritative source ('registration authority') for each metadata element exists.</li> <li>iii. <b>One entry/update:</b> Minimize errors by entering once and updating in one place.</li> <li>iv. <b>Standards variations:</b> Ensure that variations from standards are tightly managed/approved, documented and visible.</li> </ul>  |
| <b>Relationship to Statistical Business Processes</b> | <ul style="list-style-type: none"> <li>i. <b>Integrity:</b> Make metadata-related work an integral part of business processes across the organization.</li> <li>ii. <b>Matching metadata:</b> Ensure that metadata presented to the end-users match the metadata that drove the business process or were created during the process.</li> <li>iii. <b>Describe flow:</b> Describe metadata flows within and between statistical business processes (alongside data flows and business logic).</li> <li>iv. <b>Capture at source:</b> Capture metadata at their source, preferably automatically as a bi-product of other processes.</li> <li>v. <b>Exchange and use:</b> Exchange metadata and use them for informing both computer based processes and human interpretation. The infrastructure for exchange of data and associated metadata should be based on loosely coupled components, with a choice of standard exchange languages, such as XML.</li> </ul> |
| <b>Users</b>  | <ul style="list-style-type: none"> <li>i. <b>Identify users:</b> Ensure that users are clearly identified for all metadata processes, and that all metadata capturing will create value for them.</li> <li>ii. <b>Different formats:</b> The diversity of metadata is recognised and there are different views corresponding to the different uses of the data. Different users require different levels of detail. Metadata appear in different formats depending on the processes and goals for which they are produced and used.</li> <li>iii. <b>Availability:</b> Ensure that metadata are readily available and useable in the context of the users' information needs (whether an internal or external user).</li> </ul>  |



## 7. Corporate governance models for metadata management

### General considerations

It is not sensible to prescribe an ideal model for corporate governance of metadata. Every statistical organization works under different legislation, organizational arrangements, organization culture, business rules and levels of autonomy with respect to other public sector agencies.

This section looks at good practices for governance. Each statistical organization implementing a metadata management strategy will evaluate its own objectives, strategies and organizational arrangements. There is value in first considering the experience generated by organizations that have already done this.

The MetaNet project, sponsored by the European Commission, included a working group on adoption issues for statistical metadata systems. This group conducted a survey of statistical organizations in 2003, the results of which are referred to in this section, particularly those concerning barriers and organizational issues - both of which are relevant to governance.

### Lessons for good corporate governance of metadata

The following are examples of the lessons for corporate governance of data and metadata management based on experiences of statistical organizations in the implementation of a metadata management strategy:

- i. Senior managers, including the Chief Statistician, should be closely involved in developing the vision, formulating policy, approving SMS development plans and evaluating progress.
- ii. The roles and accountability of all organizational units with respect to metadata should be clear. Subject-matter areas are responsible for the creation, maintenance, re-use, and approval for dissemination of all data and metadata content for their statistical domain. A 'corporate data management unit' could be responsible for providing client support, developing and maintaining infrastructure and providing training.
- iii. The organization should adopt an information management culture. All staff must understand their responsibility to work towards achieving statistical integration, comparability of statistics across surveys and time, and to reuse statistical metadata as appropriate. These goals are achieved by adherence to the metadata management principles.
- iv. Specialists, such as business and systems analysts, information technology architects and statistical standards experts, are more likely than others to come across new opportunities for advancing better metadata integration, so a particular focus is needed on working with these groups.
- v. Make sure that your organization has an endorsed metadata strategy, including a global architecture and implementation plan, and that this strategy is integrated into broader corporate plans and strategies.
- vi. Either commit yourself to a metadata project – or don't let it happen. Lukewarm enthusiasm is the last thing a metadata project needs.
- vii. There is often scepticism in the organization against metadata projects. Metadata projects are usually strategic projects for the organization. If they should be carried out at all, managers across different levels and parts of the organization must all be committed to the project.
- viii. Metadata projects are often more abstract, complex and difficult to manage than other types of projects. These characteristics need to be recognized in project plans and the importance of communication with the rest of the organization about the project cannot be overstated.
- ix. Learn from failures and successes in other statistical organizations. Benchmarking and international cooperation are always useful. The case studies referred to in Section 8 are very valuable in this respect.

- x. Systematically use metadata systems for capturing and organizing tacit knowledge of individuals in order to make it available to the organization as a whole and to external users of statistics.

## Risk management

Realizing there are potential barriers is an important part of the management and governance of metadata projects. Appropriate risk mitigation actions are a significant component of project governance.

This subsection explores some of the potential barriers to the adoption of metadata solutions - technical, organizational and human - that were identified by the MetaNet survey referred to above. This survey included questions to identify and assess the importance of potential problems, as well as to go into more detail concerning the different aspects of human related issues.

## Important challenges to introducing metadata systems

The respondents were asked to answer the following question: "For each aspect of metadata please indicate what in your view poses the greatest challenge to the introduction or use of statistical metadata systems in your organization".

The result of this for all organizations was the following.

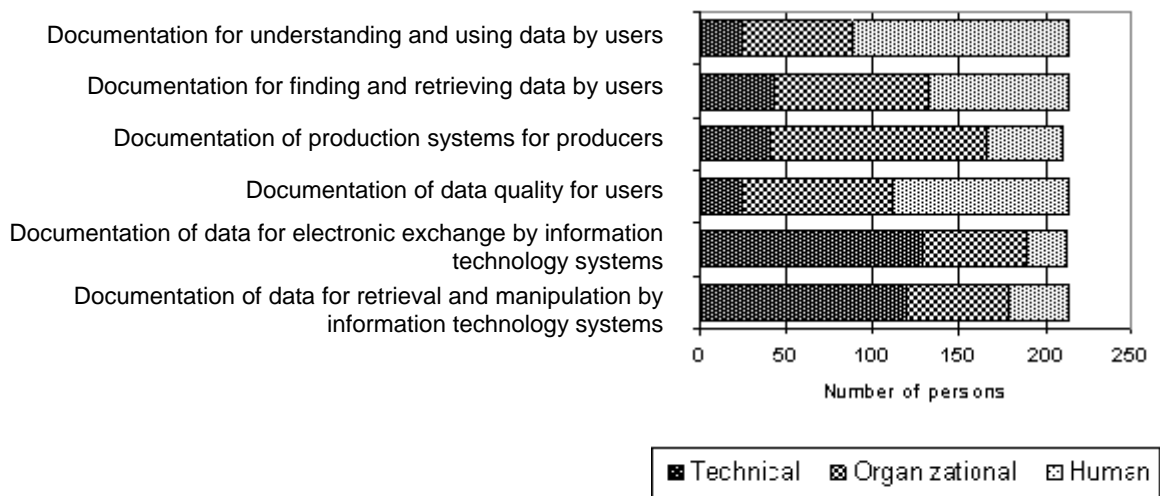


Figure 5 Perceived challenges to introducing or using SMS

The greatest challenges in relation to documentation and retrieval of data are considered to be partly organizational and partly human. However, technical challenges are the most important in relation to documenting data for exchange and retrieval by IT systems.

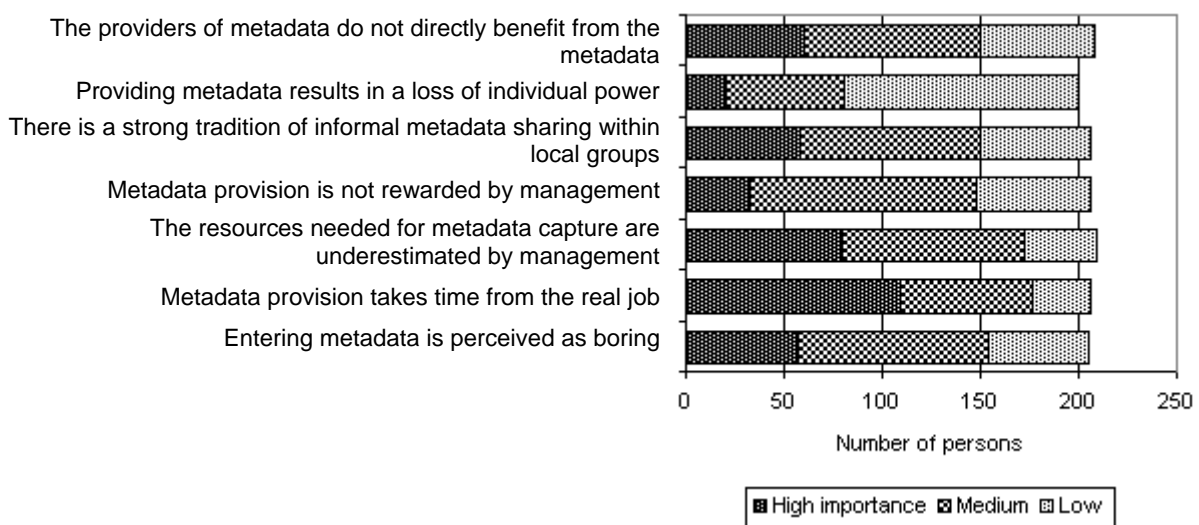
## Human issues in relation to adoption of metadata systems

The MetaNet working group reported that the human factor is fundamental to the successful adoption of metadata systems, and a number of challenges were identified.

There might be a substantial gap between some of the more theoretical and abstract contributions on metadata versus what are considered applicable by practitioners within statistical organizations. Some

subject-matter specialists tend to delve into one specific area and not take into account the long term perspective for documentation. Motivation for general metadata solutions might therefore be low. There is a need to acquire input and feedback from subject-matter specialists from different areas regarding metadata/data concepts and methods in order to come up with viable common standards and methods for metadata. However, the viability of metadata solutions requires the motivation and commitment of metadata providers. Given that an inability to engage this community constitutes a major concern for many statistical agencies, it is important that underlying human barriers are fully understood if they are to be addressed in future.

The respondents in the survey were asked to identify the most significant barriers to the provision of effective metadata within their own organization, as far as human issues are concerned.



**Figure 6 Perceived barriers to the provision of effective metadata**

The result indicates that there is no clear consensus on what the main barriers are. However, the majority think that loss of individual power as a result of providing metadata, something frequently perceived as a significant deterrent, is not significant.

Other possible issues that have often been suggested, namely that metadata provision is boring and does not benefit the provider directly were largely identified as being of only medium importance. Of greater significance was the belief that time spent providing metadata detracts from the real job. This suggests that the importance afforded metadata creation is low and that this activity will inevitably suffer at the expense of traditional work aspects. Moreover, over 25% think that management underestimates resources needed for metadata capture. If the barriers to effective metadata provision are to be overcome, the activity must be given greater importance. This demands not only the education and active involvement of would-be providers, but also increased management awareness and support.

### Organizational issues

The MetaNet report says: "It is often emphasized that it is necessary to ensure the commitment of top management in order to succeed in putting in place metadata solutions". In addition, organizational issues might be critical when planning and implementing metadata strategies and applications, even if these are not often put high on the agenda for meetings discussing metadata. Questions, such as what type of staff should be involved and to what degree there should be a central unit and which tasks such a unit should cover, must be discussed. The survey tried to uncover some of these issues.

However, three fundamental issues need reflection in order to address organizational issues:

- i. It is important to reach a common understanding within the organization of what metadata are and what their functions are. The specification of these functions will have implications on how projects should be designed and organized.
- ii. Organization of tasks related to metadata should be based on the information management strategy for the organization. One reason for the failure of specific metadata projects can be that they are not anchored in a more global view of the information architecture.
- iii. In order to sell the need for basic changes in technology or organization to improve data/metadata management, it is necessary to present the benefits and the proposed solutions in an understandable way, possibly based on practical experiences acquired in other organizations. Management is not prone to take decisions involving risk to continuity of production. Once again, management might not be a barrier, but the limiting factor might be the experts' ability to come up with convincing and practical proposals related to metadata. Proposals that reach too far and have a too long a time perspective will have difficulties as management will normally ask for results within a short time frame.

### The degree of central coordination

One might assume that a central coordinating unit is a signal that metadata/documentation is taken seriously and that there is a relatively high level of horizontal coordination. According to the responses to the MetaNet survey presented in the table below, only three statistical organizations reported having a strong central coordinating unit. The majority of them had a coordinating unit with limited tasks. Data archives apparently have a stronger central coordination of metadata. Even though data archives are often small organizations will more limited tasks compared to statistical organizations, they are information management experts and recognize the crucial role played by metadata.

	All	Statistical organizations	Data archives	Other
A. Strong central coordinating unit -	9	3	5	1
B. Coordinating unit with limited tasks - decentralized organization	20	15	2	3
C. No coordinating unit - distributed organization	5	3	0	2
D. Other/no answer	4	1	1	2
Total	38	22	8	8

### Tasks of a coordinating unit

The contact persons were asked to indicate the tasks allocated to any coordinating unit. It is interesting to note that for a majority of both the statistical and other organizations having some central coordination, an important task for this unit was to develop common systems and solutions.

These coordinating units also have an important role to play in developing common terminology and standards and to ensure general coordination and information in this field of work. Supervision and training apparently is not an important task of many units.

	All	Statistical organizations	Other
General coordination and information	20	10	10
Developing common terminology and standards	22	13	9
Developing common systems and solutions	23	14	9
Supervision and training	13	8	5
Other/not specified	2	2	0

### The involvement of different types of specialists

Information technology specialists appear in most organizations to have central positions in relation to planning and development of metadata systems and solutions, which is not surprising due to the traditional importance of metadata in computer based systems. Also, in most organizations, specialists in statistical methodology have a central role in this area. However, it is perhaps somewhat worrying that management and subject-matter specialists are to a lesser degree involved.

Expertise/Specialist	Statistical organizations		
	Central involvement	Partly involved	Not involved / not relevant
Information technology specialists	15	5	1
Management	6	11	4
Statistical methodology specialists	12	8	1
Subject-matter specialists	4	14	3

### Cooperation with other organizations

Metadata is a field of work where one should expect a large degree of cooperation with other organizations in order to ensure harmonization and exchange of best practice. The survey confirms this. A large majority of statistical organizations foresee cooperation with colleagues in other countries in this field, whereas many also see a possibility for cooperation with consultants and vendors of information technology systems. Cooperation with international organizations is also foreseen. Many statistical organizations foresee cooperation with data archives/documentation centres.

	Statistical organizations		
	Absolutely	Possibly	Not planned
Information technology system vendors/consultants	2	13	4
Other stat. org. in own country	6	5	8
Other stat. org. in other countries	14	4	1
International organizations	11	7	0
Data archives / documentation centres	5	9	5

Many large statistical organizations are searching for efficient models for handling data and metadata in an integrated way throughout the production process. Decentralization of technology, in some cases also leading to loss of central documentation of files and processes, has in many organizations made it even more important to find ways and means for coordinating documentation across the organization.

Thus it is useful to look more into the experiences of different organizational models in order to achieve common and efficient metadata solutions.

## 8. Case studies and experiences

Case studies about current metadata management systems and processes are being collected from statistical offices for publishing within the Common Metadata Framework Part D. They are published and maintained using wiki software that allows statistical offices to visit and update their own information and others to discuss it.

The currently available case studies can be accessed at [www.unece.org/stats/metis/wiki](http://www.unece.org/stats/metis/wiki), and are structured under the following headings:

1. Introduction;
2. Statistical metadata systems and the statistical business process;
3. Statistical metadata in each phase of the statistical business process;
4. Systems and design issues;
5. Organizational and workplace culture issues;
6. Lessons learned;
7. Attachments and links.

At the time of writing, the case studies from the following countries and international organizations are available: Australia, Austria, Canada, Croatia, Czech Republic, Germany, New Zealand, Norway, Portugal, Slovenia, South Africa, Sweden, and the United Nations Industrial Development Organization (UNIDO). Case studies by other organizations are envisaged.

Since there is no blueprint for design and implementation of a national SMS, the case studies represent a valuable source of information on practical SMS approaches and experiences. The content clearly demonstrates a shift in importance away from the historical focus on computer systems and design (metadata needed for building of statistical databases, data warehouses etc) to the functions and benefits of metadata across the statistical business process cycle. Furthermore, it demonstrates that statistical offices put more weight on links between the statistical information system and the SMS.

Some of the lessons learned by organizations that have contributed case studies to date are provided below.

### **Australian Bureau of Statistics**

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.
2. All high level organizational units need to be engaged by the metadata management program and have defined responsibilities in relation to it.
3. It has 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.
4. In addition to developing and deploying infrastructure, a metadata management project should be understood, and managed, as a "cultural change" initiative for an organization. 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 that are used to operating in a more autonomous and self contained and often a less structured manner.
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.

6. In addition to allowing sufficient time and resources for the business analysis, design and development process it is crucial there are sufficient resources for:
  - 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 information technology environment, evolve over time;
  - co-ordinating and promoting "outcome realisation" from the investment.
7. Business areas must be able to engage with implementation processes.
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.
10. "Statistical metadata management" is increasingly expected to interoperate with metadata management as practised in other communities (e.g. geospatial, academic/research) and sectors (e.g. use of XBRL (eXtensible Business Reporting Language) 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.

An emerging lesson, also, is that while Service Oriented Architecture (SOA) offers a lot of opportunities and potential, it also comes with a lot of new complexities compared with earlier approaches. It requires new understandings and a new mindset from those developers who are being asked to take up, and interact with, the available services as well as requiring the same from the business analysts and programmers within the team responsible for providing the metadata repositories and services. It can make the overall environment much more complicated in some ways (e.g. services are calling services that call services etc and then somewhere at a low level a service is updated and everything needs to be configured appropriately to allow proper testing of that change). Implementing SOA in environments that include a lot of "legacy" processing systems that are not enabled for the new architectural directions is particularly challenging. A highly successful example of implementing an SOA based metadata management environment would be of very high value as a case study for the Australian Bureau of Statistics.

### **Central Bureau of Statistics, Croatia**

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 national statistical organizations 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 item 1).

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 the private sector and this needs no further comment. The CROMETA project started with 16 people (more or less involved) and ended with 3.

### **Czech Statistical Office**

1. SMS strategy in terms of contents and methodology should be fully in the responsibility of the statistical office;
2. SMS design and implementation should be organized in multidisciplinary working teams;
3. Design and implementation of SMS must be managed and systematically monitored by the top management;
4. It is necessary to persistently obey the SMS core principles and to maintain a positive motivation of a wide range of participating experts (and professions);
5. Consistent co-ordination of time-scheduled workloads in the SMS project, the statistical information system redesign project and current activities of the office;
6. 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 a qualified estimate of man-hours. It is important to use all potential sources of funding (external and internal sources);
7. Financial costs of the operational running of the SMS should be covered from the office budget.

### **German Federal Statistical Office**

1. Metadata management is a communication challenge. We found two issues were particularly difficult to communicate:
  - Metadata management is tricky. Statistical data are inherently volatile. For any given data, an endless number of transformations are possible producing an endless amount of metadata. With the distribution of modern information technology 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.
2. As we are faced with multiple stakeholders, several isolated decisions taken by governing committees and a variety of information technology systems in place; it would probably be useful to develop a metadata strategy. Such a strategy might help the organization to focus on important projects and provide a coordinated approach ensuring that systems are able to interact. Distributing the energy of an organization 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.

3. The advantages and disadvantages of a metadata model can often only be properly evaluated once an information technology system is in place. It is therefore important to learn from evaluations of existing systems.
4. 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.
5. 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 our metadata team to understand the subject of metadata management. The development of information technology 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.

### **Statistical Office of the Republic of Slovenia**

1. Participation in pilot projects enables less experienced employees to gain the experience necessary for independent work. Due to the similarity of statistical processes, it is very important for IT personnel to gain experience from other statistical offices, which is ensured through participation in expert conferences and bilateral cooperation with foreign offices.
2. Metadata usage in production tools has shown many opportunities but also a lot of new challenges.
3. In 2009 within total quality management priorities will cover:
  - a. activities to promote the cooperation of enterprises in data reporting;
  - b. consistent monitoring and analysis of response burden;
  - c. implementation of the process of standardising questionnaires in statistical surveys;
  - d. preparation of documentation on the consistency and coherence of statistical data from various surveys;
  - e. preparation of internal rules and procedures on revisions to published statistical data;
  - f. establishing internal rules and procedures for keeping documentation for the time series breaks.
4. 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.
5. 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.
6. 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.
7. Problems encountered:
  - a. Complexity of the project / relatively short time for implementation. The project was very complex and ambitious. The information system covers the metadata management system, the

entire process of flexible and metadata driven statistical surveying, a sophisticated electronic reporting system, the statistical business register and the special respondent's management system. This is further complicated by the heterogeneous technological platform and all kinds of integration demands with specific statistic tools.

- b. User specifications. The consultant prepared technical documentation during the analysis and design phases, which described the future information system; however, the technical documentation in most cases was not transparent enough for users to understand the proposed system in detail before it was built. The main problem was “translation” from “statistical language” to “IT language”, which caused many misunderstandings.
  - c. Organizational adaptation of SORS. An information system like that newly developed, inevitably 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 organization of data processing has been adapted to implement the metadata driven system successfully in its processes. Implementation of electronic reporting alone demands an entirely different "service and customer-centric" organization of the data collection process.
  - d. Organization of user testing and bug repair. The user testing and bug repair process on some modules was not organized efficiently enough, which meant that the user testing and bug repair phase was not completed on time.
8. Some recommendations for the future: We highly recommend to perform smaller projects with strong project leadership; 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 consuming and all the details of the data process should be described and explained.

### **Statistics Austria**

1. It is not a new discovery that the subject of statistical metadata is an extremely complex one. Even now, almost three and a half decades after Bo Sundgren first used the term, 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.
2. 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.
3. 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 their 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.
4. 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).
5. 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.

6. 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.
7. 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.
8. 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.
9. With regard to 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.
10. 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.
11. 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.
12. 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.
13. 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.

## Statistics Canada

1. 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.
2. All of these plans and priorities for the further development and use of the Integrated Meta-database critically depend 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 Integrated Meta-database 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.
3. 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 Integrated Meta-database project are all activities that have improved the level of awareness and understanding regarding the use and importance of metadata at Statistics Canada.

## Statistics 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 organizational buy-in, the communication of the organization metadata principles and associated model is something that needs some strong consideration.
2. Everyone 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 organization - we have agreed on MetaNet, supplemented with SDMX (Statistical Data and Metadata Exchange). As Statisticians we love frameworks so having one makes life a lot easier. You could argue that the framework is irrelevant but it is 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 are 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 first principles, i.e. what is the concept behind the data item. Surprisingly it might be difficult for some subject matter areas to identify these first principles easily, particularly if the collection has been in existence for some time.
6. Some metadata are better than no metadata - as long as they are 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 a 'black market' running counter to the organizational 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 service-oriented architecture 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 service-oriented architecture concepts and application are very difficult to recruit, and equally difficult to grow.
11. The move from 'silo systems' to a “Business model Transformation Strategy” 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 organization, is very useful. This forum provides an environment in which standards can be discussed and agreed and the Committee can take on the role of the 'authority to answer to' if need be.
13. A 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 business process model 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.

### **Statistics 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 information technology-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 instead 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.

### **Statistics 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.
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 information technology 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. Training must be given 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, 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, 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.

**Statistics Sweden**

1. Involve users at an early stage in development. It makes them feel a part of the development;
2. Inform a lot;
3. Make some sort of prototype at an early stage, it is very difficult and abstract for users to describe use cases for a system with this complexity;
4. Decide what process metadata to collect and how.



## Glossary of terms and abbreviations

### Corporate metadata repository

A database system that stores metadata records for an organization or group of organizations.

### Metadata

Data that define and describe other data. Statistical metadata are defined as data about statistical data, and comprise data and other documentation that describe objects in a formalised way. (Source – SDMX Metadata Common Vocabulary, January 2009 version)

### MetaNet

MetaNet was created as a network of excellence to harmonise and synthesise statistical metadata developments. It started in November 2000 and finished at the end of July 2003. See their website at:

<http://www.epros.ed.ac.uk/metanet/index.html>.

### SDMX

A set of technical standards and content-oriented guidelines, together with an information technology architecture and tools, to be used for the efficient exchange and sharing of statistical data and metadata. (Source – SDMX Metadata Common Vocabulary, January 2009 version)

In February 2008, the United Nations Statistical Commission recognized SDMX as the preferred standard for the exchange and sharing of data and metadata, and encouraged further implementations by national and international statistical organizations. See the SDMX website at:

<http://www.sdmx.org/>.

### Statistical metadata system (SMS)

A data processing system that uses, stores and produces statistical metadata. (Source – SDMX Metadata Common Vocabulary, January 2009 version)

### Statistical information system

The processes and resources used to produce statistical information.

### Statistical meta-information system

System which uses and produces statistical metadata and which fulfils its tasks by means of functions like "statistical metadata collection", "statistical metadata processing", "statistical metadata storage" and "statistical metadata dissemination". A meta-information system may be active or passive. An active meta-information system is physically integrated with the information system containing the data that the metadata in the meta-information system inform about. A passive meta-information system contains only references to data, not the data themselves (Source - UNECE / United Nation Statistical Commission, "Guidelines for the Modelling of Statistical Data and Metadata", Conference of European Statisticians Methodological Material, Geneva, 1995).

### Senior management

The highest level of management in an organization, responsible for ensuring the organization meets its goals efficiently and effectively: May also be referred to as 'Executive' or 'Top' management.

### Statistical organization

A producer of official statistics. (Source – SDMX Metadata Common Vocabulary, January 2009 version)

### UNECE

United Nations Economic Commission for Europe

### XML

Extensible Mark-up Language – a mark-up language primarily used to facilitate the sharing of data across different systems, either within or between organizations.

## References

- ABS, *Strategy for End-to-End Management of ABS Metadata*, (2003)
- AMRADS, EU project, *Final Report of the Working Group on Metadata*, (2003)
- Booleman, M, Statistics Netherlands, *The Dutch Metadata Model*, (2004)
- Cluster of Systems of Metadata for Official Statistics (COSMOS), <http://www.epros.ed.ac.uk/cosmos/>
- Czech Statistical Office, *Conception of the Statistical Metadata System in the Czech Statistical Office*, (2004)
- Czech Statistical Office, *Global Architecture of the Statistical Meta-information System in the Czech Statistical Office*, (2005)
- Eurostat, *Eurostat Metadata: Architecture and Strategy*, (2003)
- SDMX, *Metadata Common Vocabulary*, (2009), [www.sdmx.org](http://www.sdmx.org)
- SDMX, *Content-Oriented Guidelines*, (2009), [www.sdmx.org](http://www.sdmx.org)
- Hustoft, A, Statistics Norway, *Development of a Variables Documentation System in Statistics Norway*, (2004) [http://www.ssb.no/english/subjects/00/doc\\_200713\\_en/doc\\_200713\\_en.pdf](http://www.ssb.no/english/subjects/00/doc_200713_en/doc_200713_en.pdf)
- Lindblom, H, Statistics Sweden, *The Metadata System at Statistics Sweden in an International Perspective*, (2004)
- Meliskova, J, AMRADS project, *External Users of Metadata: a Challenge for National Statistical Offices*, (2003)
- OECD, *Data and Metadata Reporting and Presentation Handbook*, (2007), <http://www.oecd.org/dataoecd/46/17/37671574.pdf>
- OECD, *Management of Statistical Metadata at the OECD*, (2006), <http://www.oecd.org/dataoecd/26/33/33869551.pdf>
- Sundgren, B, Statistics Sweden, *Advice Concerning the Strategic Decisions that have to be made by a Statistical Office when Developing and Implementing a Metadata Structure*, (2002)
- Sundgren, B, Statistics Sweden, *Statistical System: some fundamentals*, (2004)
- Sundgren, B, Statistics Sweden, *Designing and Managing Infrastructures in Statistical Organization*, (2004)
- UNECE / Eurostat / OECD, *Work Sessions on Statistical Metadata*, <http://www.unece.org/stats/archive/04.01d.e.htm>
- UNECE, *Guidelines for Statistical Metadata on the Internet*, (2000) <http://www.unece.org/stats/publications/metadata.pdf>
- United States Bureau of Labor Statistics, *Creation and Use of Metadata by Survey Methodologists*, (2002)
- Zeila, K, Statistics Latvia, *Metadata Driven Integrated Statistical Data Management System*, (2004) <http://www.unece.org/stats/documents/ces/ac.61/2004/wp.35.s.e.pdf>