BiH Metadata

Metadata users and governance

Carlo Vaccari (Istat, UniCam , ISF , MZ , ...)
vaccaricarlo@gmail.com
http://vaccaricarlo.wordpress.com

Banja Luka June 2012

SMS definition

Metadata: “data about data” or “data that define and describe other data”

Statistical metadata are “data about statistical data”

(definitions from the 2009 edition of the SDMX Metadata Common Vocabulary)

Statistical Metadata System (SMS): “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.

SMS functions

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

SMS functions 2

6. Improving the quality of statistical data. Quality = f(relevance, completeness, comparability, accuracy, timeliness, accessibility…)
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 and with administrative sources
10. Disseminating statistical information to end users: they need reliable metadata for interpretation of data

SMS functions 3

11. Improving integration between national and international organizations, in order to make statistical information more comparable

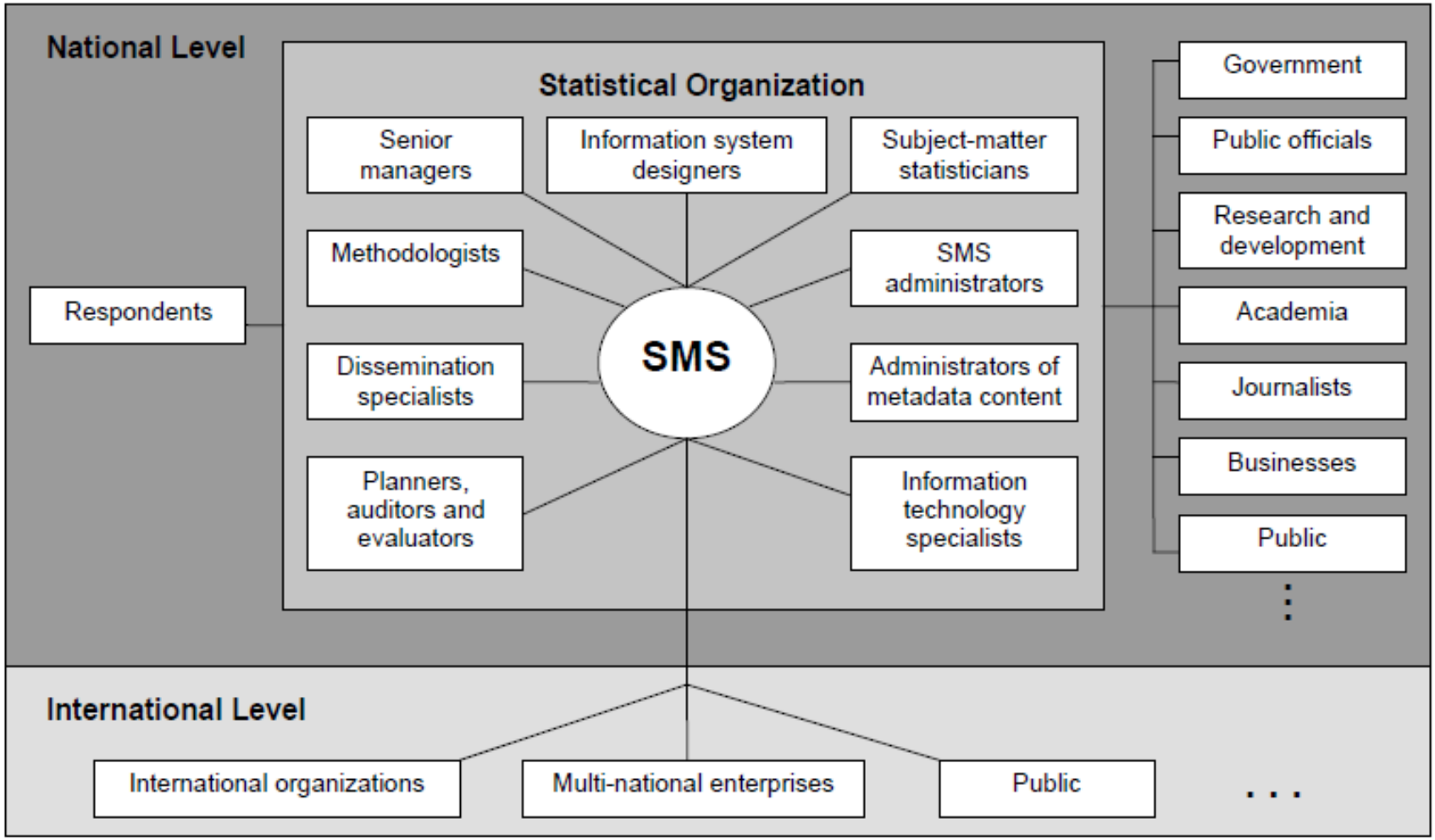
12. Developing a knowledge base on the processes of statistical information systems, to share knowledge among staff and to minimize the risks related to knowledge loss

13. Improving administration of statistical information systems: administration of responsibilities, compliance with legislation, performance and user satisfaction

14. Facilitating the evaluation of costs and revenues for NSIs

15. Unifying statistical terminology for better communication and understanding between managers, designers, subject-matter statisticians, methodologists, respondents and users

Metadata users



Metadata users

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

Metadata users

- Outside (national):
  - Respondents: respondents are important users of metadata, 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: government institutions, political decision makers, researchers, public officials, archivists, academics, librarians, journalists, businesses and the general public

Metadata users

- Outside (international):
  - 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.

Metadata benefit for all 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 search, retrieval, 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

Metadata for Senior Managers

They need tools to facilitate the design, planning, decision-making, evaluation (cost and benefits) and administrative management of statistical information systems

Following metadata will be needed:

- End-user needs and stakeholder requirements (nat/intl level)
- Available statistical services
- External information systems related to statistical systems
- Suppliers and sources of data in statistical information systems
- Statistical production process
- Publications, calendar, copyright and dissemination issues
- Responsibilities inside the statistical organization
- Costs and revenues

Metadata for IT users

Information system designers: need access to metadata from old systems, to inform the design and implementation of a new system. Need feedback about performance, costs, usage, and user satisfaction, e.g.:

- How similar systems have been designed in the past
- What observation data are available and how can be obtained
- What methods, tools and software components are available
- Detailed, up-to-date documentation of the system
- Feedback, formal or not, on system production and usage
- Experiences from similar systems
- Knowledge about methods, tools and software components
- Special evaluation studies performed on an ad hoc basis

Metadata for domain users

The SMS is a knowledge management system for the subject-matter statistician, covering many aspects, such as:

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

Metadata benefit for domain users

Subject-matter statisticians benefits:

- access to a consistent store of standard classifications, data elements, process engines that can be used in new statistical process development
- tools and links that enable the subject-matter statistician to create statistical products with a common ‘look and feel’
- a record of statistical collections and a reference point to find information on related collections (resource for new employees coming into a statistical field and for statisticians in other fields)
- standard processes, such as the registration of new data elements, which would provide a common method for the use of metadata
- current use of metadata, and any issues encountered with existing metadata; content information sharing between subject matter areas

Metadata for methodologists

Methodologists need following metadata :

- Content of statistical data (micro/macro) and associated concepts
- Quality of statistical data (relevance, accuracy, timeliness, …)
- Existing statistical tasks and surveys (questionnaires, sources, etc)
- End users and their feedback
- Requests of international organizations and related standards
- Respondents’ information systems
- Administrative data, statistical registers (population, farms etc)
- Data sources, information systems and their output databases
- Statistical classifications, nomenclatures and standards
- Statistical population, units, measurement units, time series
- Statistical methods and relevant research projects

Metadata for SMS administrator

Administrators of metadata content

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, to avoid inconsistencies and redundancies

The administrator will need a user-friendly interface, avoiding any technical skill.

The administrator will need the following metadata:

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

Metadata for SMS Administrator

SMS administrators are responsible for the technical maintenance of the SMS. They should cooperate with designers, evaluators and content administrators in solving technological issues and in development of the SMS.

The technical administrator will use 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 used by content administrator

Metadata for IT specialists

IT specialists

Metadata on the content of statistical data (classifications, units, measurement unit, time series, 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

Ideally, statistical production processes will generate metadata about their own performance, giving feedback about functioning and efficiency of metadata driven production

Metadata for respondents

Respondents and data supplier

Respondents are important partners of any statistical information system. Statistical data suppliers are often also users of statistical data: their role is becoming more important with the growing number of systems integrated and online communication possibilities

There is a growing need to harmonize methodological definitions of data and related metadata from respondents and statistical information systems. 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

Metadata for respondents

Respondents and data suppliers will require the following info:

- Metadata related to the content (definitions, terminology) of statistical data
- Security and confidentiality of microdata
- Feedback from statistical outputs
- Information about the content of statistical warehouses
- 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
- Potential interfaces between statistical information systems and respondents' information systems
- Information about software supplying data and metadata

Metadata for end-users

With the increasing 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.

Increase in the reuse of data by external websites, web services and 'mashups' → metadata must be closely coupled with data such as both web services and people can use

This heterogeneity of sources, together with methodological differences and inconsistencies of statistics disseminated via the Internet, poses difficulties for the users → there is a need for harmonization of metadata accompanying statistical information on the Internet. (International standards should play an important role)

Metadata for end-users 1

The following metadata are vital for end users :

- Metadata and data concepts and definitions, classifications, aggregations, statistical and evaluation methods, terminology, history
- Metadata about quality (e.g. explanatory notes)
- Access to microdata
- Time series
- Updating procedures and statistical revisions
- Responsibility for individual statistical outputs
- Links to other information systems both national and international
- Confidentiality
- Content related standards, both national and international;
- Outcomes from statistical analysis on users feedback

Metadata for end-users 2

The following metadata are vital for end users :

- Rules for searching, accessing and downloading statistical metadata and data from output databases
- Technological standards for extraction 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 of statistics can apply
- 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 (Open Data)

Metadata for international users

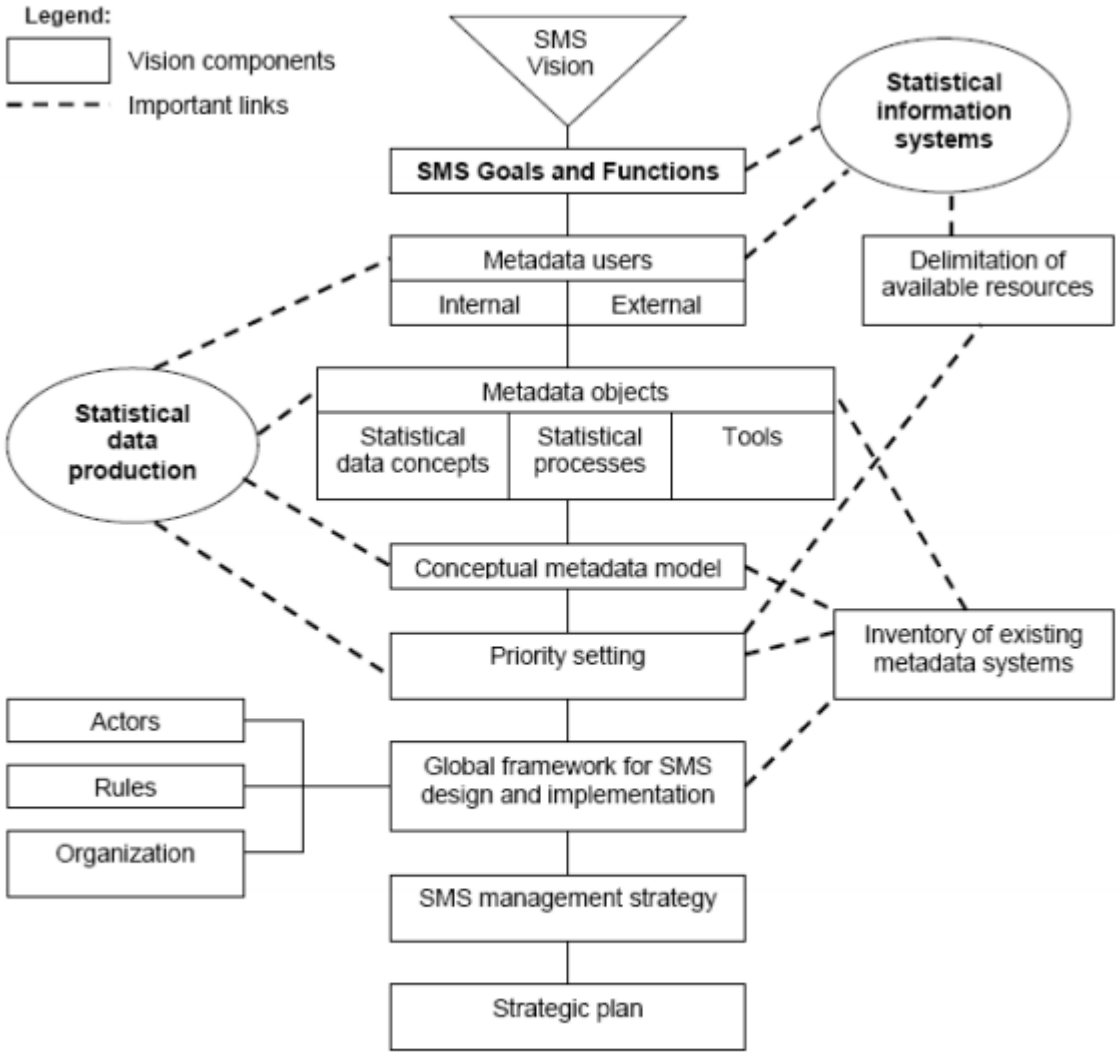
The needs of international users increasingly impact the architecture of the SMS of national statistical organizations. At the same time there are unprecedented rates of change in technologies, end user expectations, information needs and data sources, which require national statistical offices to collaborate in order to meet international user requirements

Metadata needed by international users are those needed by end users on national level plus the following information :

- Compliance 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 standardization of statistical data and metadata concepts

Metadata vision and strategy

- Vision and GSBPM



Core principles (next mission)

Metadata handling

i. Statistical business process model: Manage metadata with a focus on the overall statistical business process model (www.unece.org/stats/gsbpm).

ii. Active not passive: 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.

iii. Reuse: Reuse metadata where possible for statistical integration as well as efficiency reasons

….

SMS governance

Lessons for good corporate governance of metadata:

i. Top managers should be closely involved in developing the vision, formulating policy, approving SMS development plans and evaluating progress

ii. The roles of all organizational units should be clear. Subject-matter areas are responsible for data and metadata content for their statistical domain. A 'corporate metadata unit' could be responsible for client support, infrastructure and 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.

iv. Specialists must come across new opportunities for advancing better metadata integration.

SMS governance

v. The organization must have a metadata strategy, including a global architecture and implementation plan, and this strategy must be integrated into corporate plans and strategies.

vi. Either commit yourself to a metadata project – or don't let it happen.

vii. There is often scepticism in the organization against metadata projects. Metadata projects are usually strategic projects for the organization. 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. (communication!!)

ix. Learn from failures and successes in other statistical organizations.

x. Systematically use metadata systems for capturing and organizing tacit knowledge of individuals in order to make it available

SMS governance

Organizational issues:

i. It is important to reach a common understanding within the organization of what metadata are and what their functions are.

ii. Organization of tasks related to metadata should be based on the information management strategy for the organization. One common reason for the failure of 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. Proposals that reach too far and have a too long a time perspective will have difficulties as management will normally ask for results in a short time.

SMS governance

Degree of central coordination

The existence of 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 MetaNet survey, only few statistical organizations reported having a strong central coordinating unit. The majority of them had a coordinating unit with limited tasks.

Tasks of a coordinating unit

- developing common solutions
- developing common terminology and standards
- ensure general coordination and information in this field of work.
- training

SMS governance

Recommendations for SMS management

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 top management, 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.

SMS governance

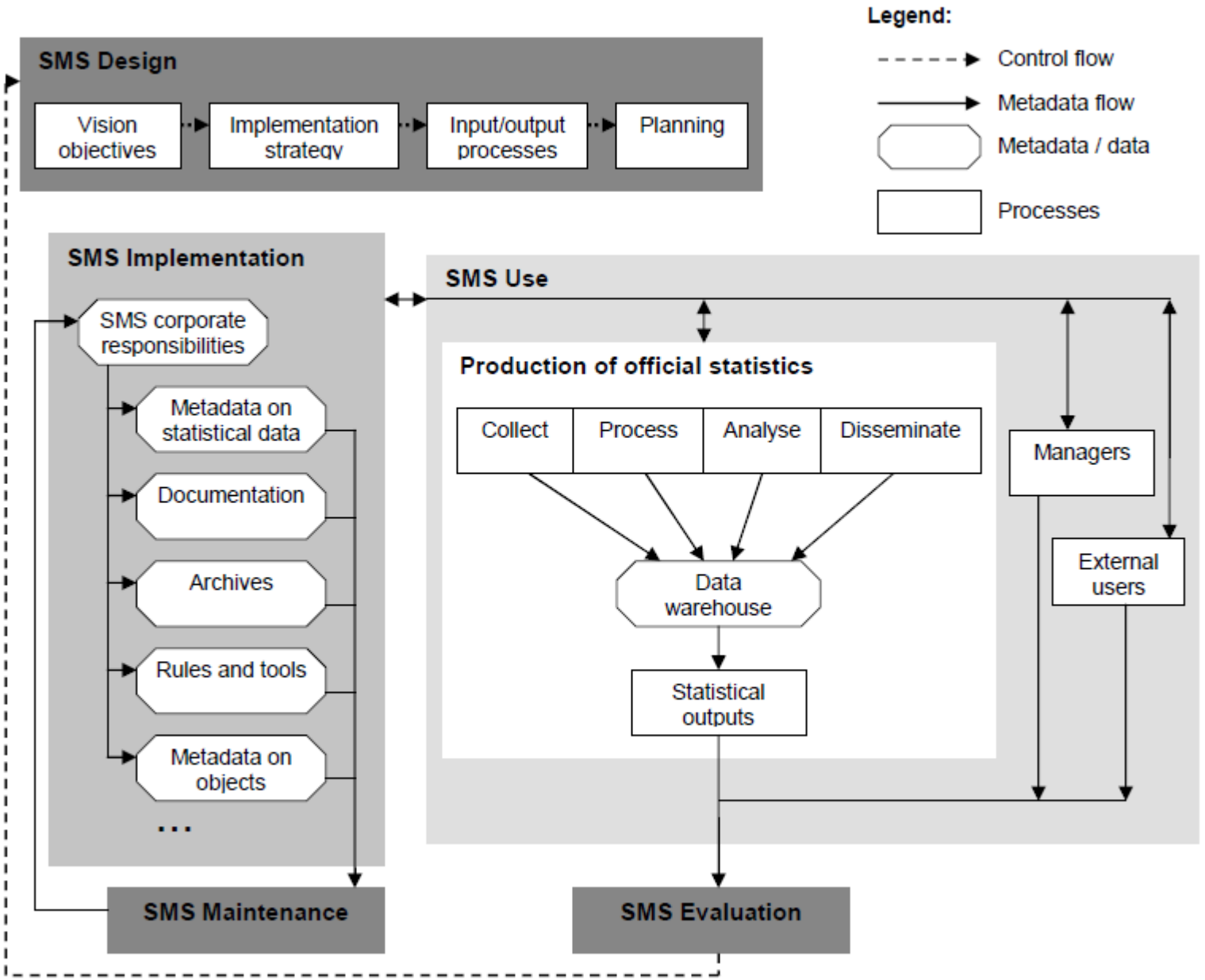
Recommendations for SMS management

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 IT specialists.

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. Many critical issues could appear. Such issues should be foreseen and reflected in the SMS vision and strategic plans as far as possible.

SMS governance



Management of SMS Design

Management of SMS design

The most important functions, tasks and activities 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, describing all processes that will work with metadata. An inventory of all such processes and existing metadata tools should be prepared in close cooperation with major stakeholders.

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.

Management of SMS Design

The most important functions, tasks and activities are as follows:

v. The impact of the corporate SMS on existing statistical production systems should be clear from the SMS vision. Any necessary re-engineering should be considered and planned at this stage.

vi. All the points of contact between the SMS metadata model and business processes, in terms of creation, update, and use activities, are described.

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.

ix. Financial requirements for implementation should be detailed.

Management of SMS Implementation

The following major activities should be considered:

i. An agreed set of definitions and terminology should be developed. Consideration of standards 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.

iii. Existing processes using statistical metadata should have been reengineered.

iv. Outsourcing 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 and the harmonization of technical administration.

Management of SMS Implementation

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.

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 staff should be recognised.

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. Training for all metadata users should be organized.

Management of SMS Use

Management of SMS use tasks:

i. Prepare, maintain and coordinate detailed plans of metadata use to ensure the required metadata quality

ii. The units responsible for statistical production should prepare the plans related to the activities dealing with the production process

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 regular feedback from users about metadata quality and the availability of metadata tools. Use also surveys

vi. SMS management should be aware of the technological environment related to the use of metadata and metadata tools.

vii. Statistical websites must be an integral part of SMS

Management of SMS Maintenance

Management of SMS maintenance tasks:

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. A detailed plan of maintenance activities should be approved

iii. Ensure timeliness and coherence of maintenance activities.

iv. The ownership of metadata, what is the 'standards' for classifications, what are the permitted variations from the 'standard' 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.

Management of SMS Maintenance

Management of SMS maintenance tasks:

vi. SMS management is responsible for definition of policies, procedures and protocols for the maintenance of the repository.

vii. Rules and guidelines should be developed for the maintenance of each metadata entity, identifying the responsible metadata owner.

viii. Preparation of rules and guidelines requires joint work with metadata owners. Methodologists are also important partners.

ix. Training of metadata owners in the rules and guidelines prepared for maintenance activities will be required.

x. SMS management should ensure updating of links between metadata in the corporate metadata repository.

xi. All maintenance functions performed by metadata administrators and metadata owners should use a standard set of metadata tools.

Management of SMS Evaluation

Management of SMS evaluation tasks:

i. Specify the major targets of SMS evaluation and prepare a plan of evaluation activities and procedures based on these.

ii. Evaluation of user satisfaction should be an ongoing part of the SMS development life cycle. The most important object of evaluation is the external user of data and metadata.

iii. Other important aspects for evaluation are cost efficiency, implementation of standards, organization of work, maintenance procedures and technological implementation.

iv. Three major forms of evaluation: (i) regular long-term evaluations (e.g. at 3 year intervals) that examine overall effectiveness of SMS; (ii) regular short-term evaluations (e.g. annually) on user satisfaction; and (iii) ad hoc evaluations as necessary.

v. Benchmarks should be established for all defined targets, agreeing on benchmarking parameters and evaluation methods.

Management of SMS Evaluation

Management of SMS evaluation tasks:

vi. 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 top management of the statistical organization on the evaluation outcomes, including prioritized recommendations. Based on the conclusions made by top management, organize improvement of and/or further development of the SMS.