Handbook on Statistical Metadata for the ESCWA region

Table of contents

Executive Summary

The 'Handbook on Statistical Metadata for the UN-ESCWA region' was prepared as part of UN-ESCWA initiative to strengthen member states knowledge in the area of metadata.

The main objective of this handbook is to provide a comprehensive review of the current situation on the work done in the area of metadata in the region and to bring to attention of national statistical offices (NSO's) in the UN-ESCWA region to issues related to the strategic meaning of statistical metadata, standards applicable to microdata level and tools that can be used for statistical metadata management at national level.

This handbook can be used as a guideline tool for NSO's to initiate or develop their ongoing activities in terms of metadata that is considered as a powerfully mechanism to communicate data with all types of users. It will also introduce a number of modern tools for data management, microdata documentation and archiving and data dissemination, that are in compliance with UN recommendations.

This handbook is made up of four parts. The first part is dedicated to strategic aspects of statistical metadata; it highlights positive elements of metadata strategies in member countries that can provide an example and guidance to other national statistical offices in the region.

Part two is on statistical standards applied, providing an overview of metadata standards applied by national statistical offices of member countries in their metadata management.

Part three is related to the statistical business process and statistical metadata; it provides an overview of metadata used at each phase of the Generic Statistical Business Process Model (GSBPM), as identified by national statistical offices of member countries.

Part four is dedicated to metadata management software tools providing practical information for NSO's where to look for useful tools and experiences, focusing mainly on experiences of Palestine, Jordan and Egypt using the Microdata Management Toolkit.

Acronyms

DC	Dublin Core
DDI	Data Documentation Initiative
ECB	European Central Bank
EUROSTAT	
ICPSR	Inter-university Consortium for Political and Social Research
IMF	International Monetary Fund
IHSN	International Household Survey Network
ISO	International Organization for Standardization
GSBPM	Generic Statistical Business Process Model
Nesstar	Metadata editor and publishing tool
NSD	Norwegian Social Science Data Services
NSO	National Statistical office
OECD	Organization for Economic Co-operation and Development
RDF	Resource Description Framework
SDMX	Statistical Data and Metadata eXchange
SDDS	Special Data Dissemination Standard
SMS	Statistical Metadata Systems
UN	United Nations
WB	World Bank
XBRL	Extensible Business reporting language
XML	Extensible Markup Language

Introduction

The aim of this handbook is to explain and define core metadata elements, initiatives and tools as defined at the international level and to present what UN-ESCWA member states have achieved to date in terms of using metadata in their statistical work. Countries experiences are highlighted in this handbook, and this is basically based on: i) the situation assessment done by the UN-ESCWA using the specific questionnaire (*Metadata case study template*) that member states completed in 2011, ii) and on the Expert Group Meeting on Data and Metadata Reporting that was organized by the UN-ESCWA in Amman, Jordan, 16-18 July 2011. Authors also used research of websites of national statistical offices, as well as their knowledge of metadata issues in order to complete this handbook.

By publishing this handbook, we plan to increase awareness of managers in national statistical offices of activities taken by various countries in the ESCWA region in building and further developing statistical metainformation systems. In particular we would like to draw their attention to the role of metadata and metainformation system in the strategic management of national statistical offices and in integrating the statistical operations.

The handbook is intended as a basis for creating a living tool on metadata exchange and facilitating exchange of experience between national statistical offices. The further updates of this handbook will aim at bringing in more of national experiences and treating specific needs of countries in the ESCWA region.

What is metadata?

Metadata describes other items of content. A traditional example of metadata is represented by are catalogues and card registries that may describe books in libraries, physical assets of the institutions, individual persons, etc. These were largely computerized and converted to database, but metadata comprised in them still have their very original purpose. Metadata for an image may include description of the content, title, but also technical parameters that is the size of picture, the color depth, format, the digital resolution, author, date of creation and other information. A text document's metadata may contain information about authorship, copyright, language, length, date of creation/modification, a short summary and some technical parameters of the document. Metadata are often attached also to software and other products providing detailed information needed for maintenance and adaptation to the user's needs. Web pages often comprise metadata in a form of meta-tags that facilitate indexing by search engines.

What is statistical metadata?[1]

Statistical metadata are data which are needed for proper production and usage of statistical data. They describe statistical data and –to some extent- processes and tools involved in the production and usage of statistical data. Expressed briefly, statistical metadata are data about statistical data. (Source: ISO/IEC FDIS 11179-1"Information technology − Metadata registries – Part 1: Framework", March 2004)

---

[1] - "Guidelines for the modeling of statistical data and metadata" UNECE ,1995.
http://www.unece.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/DAM/stats/publications/metadatam odeling.pdf&t=1323105817&hash=0c9996299d5d071dbca1e1f5ed283058a214e3b8

What are the different types of statistical metadata?

While this is a frequently asked question, there is not a straightforward answer to it. We intuitively understand that there is a difference between titles in the stub and header of a statistical table, footnotes, definitions of indicators, sources. We all perhaps understand that there is a difference between metadata that define a structure, that provide a description and that are geared towards statistical methodologies.

One of the approaches presented within the Common Metadata Framework[2] results in the following disctiction between metadata types:
- Conceptual metadata;
- Process metadata;
- Quality metadata;
- Technical metadata.

Another distinction follows the viewpoint of electronic data interchange to:
- Administrative metadata;
- Structural metadata;
- Conceptual/methodological metadata;
- Footnotes.

The Common Metadata framework also distinguishes between metadata according to the `the information entity they describe[3]:
- Users (customers) 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

Later in this handbook (Par III) we will also distinguish between metadata used at various phases of the statistical survey process. Another way of distinguishing between metadata about techniques and methods are whether they describe the methodologies **ex ante** (that is how they should be applied) or **exp post** (that is how they were applied.

As a conclusion, it is important to take various types of metadata into consideration when designing metadata strategies, metainformation systems and databases. However, there is not a unique way of classifying various types of metadata. All it depends on the context in which metadata are used and the information systems are being applied to.

2

http://www1.unece.org/stat/platform/display/metis/3.+Metadata+in+each+phase+of+the+statistical+business+process+%28Netherlands%29

[3] http://www1.unece.org/stat/platform/display/metis/Part+A+-+Statistical+Metadata+in+a+Corporate+Context

Part I: Strategic aspects of statistical metadata

Statistical Metadata System

Statistical metadata systems (SMS) play a fundamental role in statistical organizations. Such systems comprise the people, processes and technology used to manage statistical metadata. A traditional view of statistical metadata emphasized their descriptive role when data were disseminated to the users. However, in modern statistical organizations metadata accompany data throughout the statistical survey cycle, and allow to better exchange data between various units and to integrate processing of statistical data. It means that statistical metadata and statistical metainformation systems have a key role in integrating individual statistical areas. This brings new opportunities to the strategic management of statistical offices, because advanced integration bring with it new architecture of statistical systems orientated on processes, it allows getting out of silos and develop statistics in cross-cutting areas better covering the image of the complexity of current societies.

As a particular example, integrating demographic, social, economic and environmental data is possible thanks to the ability of metainformation systems to record concepts and definitions (so statisticians can verify their cross-domain compatibility), and methods of data integration (so users can verify the relevance of results). Modern metainformation systems are able to provide the management of statistical offices with exhaustive information about statistical surveys including design, methodologies and statistical processes, but also mandate, resources, responsible persons, lessons learned, etc.

In order to achieve costs efficiency, data are often combined from a variety of sources rather than a single survey. There is an extensive use of administrative registers and records and combining data from various administrative sources and from traditional statistical surveys. Use of such combined sources of data requires detailed metadata that allow for verification of concepts, definitions and their consistency.

An anecdotal evidence of the strategic role of metadata is that most of the work on unifying the formats for exchange of statistical tables and time series[4] focused on modeling statistical metadata. These unified formats attempt to automate and facilitate exchange of statistical data between organizations and between various databases. The focus was put on the exhaustive description of statistics exchanged. Therefore, these formats comprise advanced structures for various kinds of metadata.

A statistical metadata system 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

---

[4] The authors refer to the two decades of work on standardising electronic exchange of statistical data originating from the Generic Statistical Message GesMes and resulting in the Statistical Data and Metadata eXchange initiative (SDMX)

organization's strategic direction. Therefore, it is important that the work on metainformation systems within a statistical office involves not only metadata specialists, but also managers at all levels.

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 a number of national and international statistical organizations all over the world. 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.

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.

What is the role of the statistical metadata system?[5]

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 for the statistical viewpoint (needs, concepts, definitions, methods and techniques, but also mandates, resources and costs, roles and responsibilities and lessons learned).

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

---

[5] Common Metadata Framework Part A: Statistical Metadata in a Corporate Context: see
http://www1.unece.org/stat/platform/display/metis/Part+A+-+Statistical+Metadata+in+a+Corporate+Context

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.

Users of the statistical metadata system[6]

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 the following figure

---

[6] [6] Common Metadata Framework Part A: Statistical Metadata in a Corporate Context: see
http://www1.unece.org/stat/platform/display/metis/Part+A+-+Statistical+Metadata+in+a+Corporate+Context
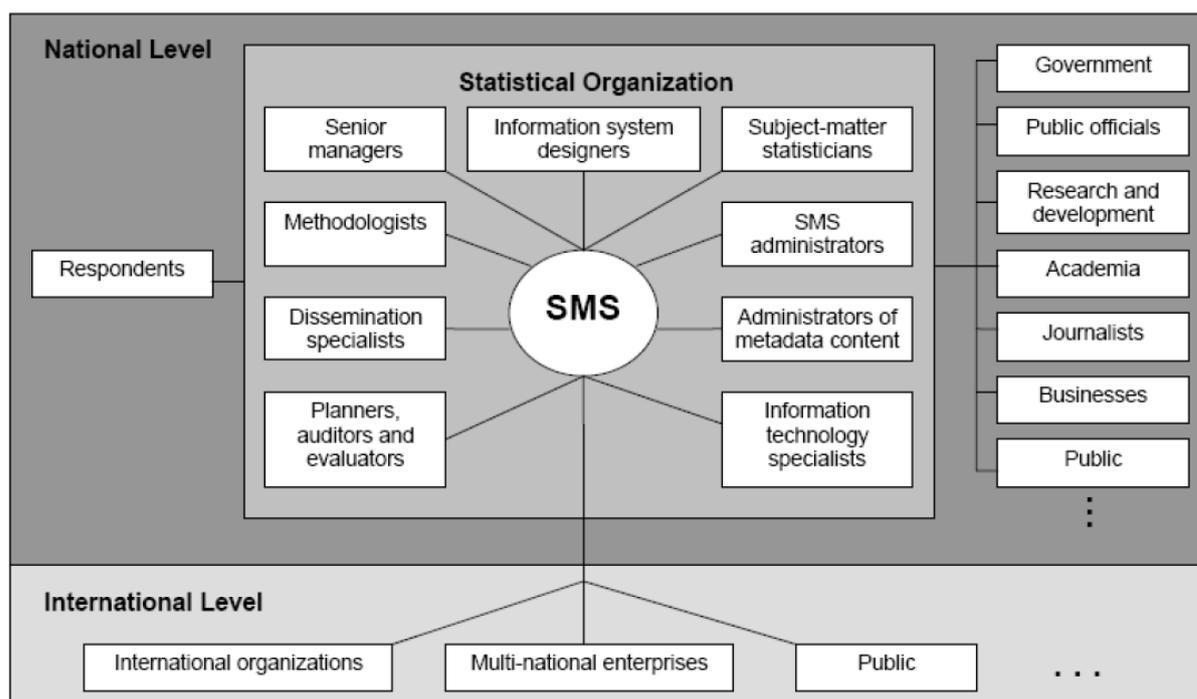
Taking into account the interest of these different groups of users and the benefits that each group can expect of the SMS a lot of effort should be made by the NSOs to design SMS that satisfy these needs.

More and more countries over the World are pulled toward developing strategies to implement SMS according to their needs and to available resources. Several interesting countries experiences can be found on UNECE website.[7]

Several countries of Middle East region have started being interested in the development of Metadata. The situation is different from one country to another; this situation will be explored in the following section.

Assessment of the current situation regarding the adoption of metadata strategies in UN-ESCWA member countries

The UN-ESCWA is made up of 14 member countries, namely: Egypt, Bahrain, Iraq, Jordan, Kuwait, Lebanon, Oman, Palestine, Qatar, Saudi Arabia, Sudan, Syria, United Arab Emirates and Yemen. Checking national statistics offices' websites and the questionnaires provided to UN-ESCWA, 6 out of 14 countries had strategies regarding implementing or planning to use any metadata initiative.

The case of Bahrain

The Central Informatics Organization (CIO) of the kingdom of Bahrain has a clear vision with regard to the metadata strategy; their new statistical metadata approach consists of developing electronic statistical data warehouse that can be accessed online.

The current situation at the CIOas presented during the Expert Group Meeting in July was is as follow:

---

[7] http://www1.unece.org/stat/platform/display/metis/Metadata+Case+Studies

- Infrastructure is ready (GDN, NDS)
- High Level support (Prime Minister office)
- Metadata collected, processed, analyzed, disseminated and archived in many entities following different methodologies, definitions, classifications & Standards.
- Metadata exchange very difficult.
- Difficulty in liaise and coordinate with data producers and the Statistics office.
- Metadata reports not consistent
- No flexible report tools

The case of Egypt

The Central Agency for Public Mobilization and Statistics (CAPMAS) strategy for archiving statistical work is to start first with current statistical work and archive previous surveys at later stages. Horizontal and vertical expansions in terms of data archiving are planned to be conducted in parallel. The CAPMAS dedicated on its main webpage a specific tab for metadata. As stated by the agency itself, the metadata main objective is to meet users' needs in terms of statistical data in order to support policies formulation based on international best practices.

Metadata is then considered as a mechanism to communicate with all users without exception or differentiation between them.

For this specific purpose the CAPMAS has used a number of modern tools for data management, microdata documentation and archiving and data dissemination. Those tools were mainly in compliance with the World Bank and the International Household Survey Network (IHSN) (Microdata Management Toolkit) guidelines and recommendations regarding data production and dissemination.

Next steps for CAPMAS will be the use of SDMX to facilitate the data exchange, the documentation of surveys carried out by third-party, offer training to other government agencies producing data on the documentation and check on updates related to the metadata and data archiving.

CAPMAS's strategy as presented in July 2011 at the EGM has the following points:

- Building a statistical database and metadata database relying on international standards (SDMX, DDI, DCMI, SDDS)
- Study the international standards that support the proper use of the data and metadata and use the most suitable.
- Focus on the production and use of metadata at each stage of a project
- Provision and dissemination of data in Arabic and English
- Develop the capabilities of CAPMAS in the dissemination of data and metadata
- Dissemination of experience and knowledge in the field of production and use of statistical data and metadata to government agencies and to public and private sectors

The case of Jordan

The Department of Statistics (DOS) in Jordan has just started the metadata initiative. In collaboration with the OECD, DOS trained so far 18 employees on the use of the Microdata Management Toolkit and created working groups from both IT department and survey departments to start archiving current and previous surveys. Their plan is basically to use the Microdata Management Toolkit. They will train a member of each directorate to be a focal

point of his directorate to start documenting all his directorate's surveys. They have started with three directorates.

The case of Lebanon

The main webpage of the Central Administration of Statistics in Lebanon has a dedicated tab for metadata that basically has a limited number of definitions and concepts used in social surveys in addition to a Ms. Office Word document with regard to the definition and methodology for compiling the consumer price index. It has been mentioned, however, that this page will be updated continuously.

The case of Palestine

The Palestinian Central Bureau of Statistics (PCBS) has clear strategic objectives for improving the use and dissemination of statistical data and statistical information. Based on good practices in this specific area, the PCBS has developed a general framework for the metadata strategy implementation, taking into account national needs, to meet regional and international commitments.

The bureau classified the work on metadata as one of the bureau's priorities during the period 2009-2013 and within the important development projects which is included on the agenda of the bureau in all of its meetings to take place during the period mentioned computerization and dissemination of all groups of metadata in the bureau archives.
The bureau is currently producing metadata of raw data which is made available immediately after the completion (implementation) of the surveys and dissemination of results. The bureau studied the international frameworks and good practices in the area of producing and disseminating the metadata. A specialized team has been formed to assess the frameworks used internationally by specific criteria. Metadata is being archived and disseminated in the form of DDI format through a set of processing and disseminating of metadata ADP.
Currently, significant set of metadata had been published in both Arabic and English version (13 in Arabic and 20 in English) on a particular site linked to the PCBS website.

The case of Qatar

The National Strategy for the Development of Statistics (NSDS) in Qatar follows international standards and addresses all statistical aspects including data collection, analysis, dissemination and use, from censuses, surveys and administrative systems, as well as mechanisms for coordination and consultation.
The NSDS work builds upon various initiatives already in place, including data production initiatives for example the International Monetary Fund's General Data Dissemination System.[8]

The Qatar Statistics Authority (QSA) is releasing metadata through its publications and in QALM website along with the data. QSA is at the stage of planning to implement SDMX.
The QSA is planning to link its own website with the Qatar Information Exchange (QIX)[9] website that is a flexible statistical database portal containing micro data, which will allow users to extract, tabulate data/statistics, it also provides detailed statistics along with their metadata.

---

[8] http://www.qsa.gov.qa/Eng/NSIQ.htm
[9] http://www.qix.gov.qa/portal/page/portal/qix/subject_area/Metadata?subject_area=190

Part II: Statistical standards applied

This part will provide an overview of metadata standards in general and highlight the ones applied by national statistical offices (NSOs) of UN-ESCWA member countries. The reason for including this part in the handbook is that NSOs often ask for advice on which of the existing metadata standards is applicable in a specific situation, which is the one to take consideration in designing their own statistical metainformation system and defining metadata strategies.

The data archive community has developed a relatively large set of metadata standards. It is not easy to pinpoint a single standards that would serve a basis for formulating the metadata strategy. Each of the standards focuses on another issues of facilitating data communication between organizations and software systems, and to improve the quality of statistical documentation provided to users of data. These metadata standards provide a structured framework for organizing and disseminating information on the content and structure of statistical information. To take full advantage of the web technology, most standards are defined in the XML language. The Data Documentation Initiative specification (or DDI) is a standard dedicated to the documentation of microdata. The Dublin Core (DC) provides a standard for the documentation of related resources.

What are the metadata statistical standards used by NSO's in the UN-ESCWA region?

Dublin Core (DC) and the Dublin Core Metadata Initiative (DCMI)

The Dublin Core is a standard for metadata established by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship and practice, and it was conceived to facilitate the search for online information not only by specialists in resource description but also by non-specialists. This feature has been the starting point of the project. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio; "core" because its elements are broad and generic, usable for describing a wide range of resources.[10]

The Dublin Core standard includes two levels: Simple and Qualified. Simple Dublin Core comprises fifteen elements; Qualified Dublin Core includes three additional elements (Audience, Provenance and Rights Holder), as well as a group of element refinements (also called qualifiers) that refine the semantics of the elements in ways that may be useful in resource discovery.[11]

---

[10]:http://dublincore.org/
[11] http://dublincore.org/documents/usageguide/

The fifteen elements of the Simple Dublin Core

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description:

Content	Intellectual property	Identification
Coverage	Author / creator	Date
Description	Contributor	Format
Language	Publisher	Identifier
Relation	Rights	Type
Source		
Subject		
Title		

The main DC metadata elements characteristics are the following:

- Each element is optional
- Each element may be repeated
- The order of the elements is not important
- It is flat, non-hierarchical element structure
- It is extensible by refinement (sub-properties of existing properties) or by combination (application profiles)
- It is syntax independent (such as using HTML/XHTML, XML, RDF/XML)

A major reason behind the success of the Dublin Core metadata standard is its simplicity. From the outset it has been the goal of the designers to keep the element set as small and simple as possible to allow the standard to be used by non-specialists. The purpose of the standard is to make it easy and inexpensive to create simple descriptive records for information resources, while providing for effective retrieval of those resources on the Web or in any similar networked environment.

The Dublin Core Metadata Initiative (DCMI) is an open organization supporting innovation in metadata design and best practices, its aim is to promote widespread acceptance of metadata standards and best practices.

Data Documentation Initiative (DDI)

The Data Documentation Initiative (DDI) is an effort to establish an international XML-based standard for microdata documentation. Its aim is to provide a straightforward means to record and communicate to others all the salient characteristics of micro-datasets. The DDI specification is a major transformation of the once-familiar electronic "codebook," which retains the same set of capabilities but greatly increases the scope and rigor of the information

contained in it. The DDI metadata specification originated in the Inter-university Consortium for Political and Social Research (ICPSR), a membership-based organization with over 500 member colleges and universities around the world. It is now the project of an alliance of institutions in North America and Europe. The member institutions comprise many of the largest data producers and data archives in the world. An important goal of the initiative is to become an ISO standard. The most recent version of the DDI specification is version 3.0. Version 2.1 is however the most widespread.[12]

By creating a consistent framework for microdata documentation, the DDI has the following features:

- Interoperability: DDI-compliant documentation can be exchanged and transported seamlessly, and applications can be generically written, because the documents are homogeneous.
- Richer content: The DDI provides data analysts with broader knowledge about data content, because the DDI initiative provides a comprehensive set of elements that can describe micro-datasets as completely and as thoroughly as possible.
- Multipurpose documentation: A DDI codebook can be restructured to suit different applications, because it contains all the information necessary to produce different types of output.
- On-line analytical capability: DDI documents can be easily imported into on-line analysis systems, rendering datasets more readily usable by a wider audience. This is made possible because the DDI markup extends down to the variable level and provides a standard uniform structure and content for variables.
- Search capability: Field-specific searches across documents and studies are made possible, because each of the elements in a DDI-compliant codebook is tagged in a specific way.

Metadata Registries (ISO 11179)

The International Standard ISO/IEC 11179-1 describes the standardizing and registering of data elements to make data understandable and shareable. Data element standardization and registration as described in ISO/IEC 11179 allow the creation of a shared data environment in much less time and with much less effort than it takes for conventional data management methodologies.[13]

Statistical Data and Metadata Initiative (SDMX – ISO 17369)[14]

Focusing on time series and indicators, SDMX is the result of a joined effort from the Bank for International Settlements, the European Central Bank (ECB), EUROSTAT, the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), the United Nations (UN), and the World Bank (WB) to create an XML specification to support the exchange of aggregate data and metadata. SDMX provides three types of statistical metadata standards: standards for data formats, standards for metadata and a registry-based architecture to implement these standards and to exchange data between systems. One of the requirements of SDMX was the awareness of other metadata

---

[12] http://www.internationalsurveynetwork.org/home/index.php?q=tools/documentation/standards
[13] Source: ISO-IEC 1999, available at http://metadata-stds.org/11179-1/ISO-IEC_11179-1_1999_IS_E.pdf
[14] See http://www.sdmx.org for further reference.

specifications such as the Data Documentation Initiative (DDI). Any of the DDI metadata - which emphasizes archival metadata and micro-data, rather than aggregate data - is exchangeable in an equivalent SDMX metadata format. This ensures inter-operability of metadata across namespaces.

SDMX is a family of metadata standards rather than a unique standard. The very original part of SDMX were standards for structure of interchange messages, but these were soon completed by guides of less technical nature resulting in:

- SDMX technical standards (SDMX-EDI (GesMes TS) using UN-EDIFACT syntax and SMDX-ML using XML syntax).
- SDMX Content Oriented Guidelines.
- SDMX tools.

SDMX was developed in order to facilitate exchange of **aggregated** statistical data and metadata between international organizations and from national statistical offices to international organizations. There are many lessons that can be learned from SDMX for other areas of application. However, one should keep in mind also the limitations. A national statistical office has to look into other metadata standards, in particular for its microdata operations,. SMDX standard may be applicable in dissemination of statistics, exchanging of data between different phases, applying definitions and concepts from the content oriented guidelines, navigating users through the data focused websitesl.

Common warehouse model ISO 19504 and data warehouse

Specification for the metadata in support of exchange of data between tools, used as a mean for recording the metadata to achieve data exchange between tools. Its ISO /IEC number is 19504.
It is a Standard interfaces that can be used to enable easy interchange of warehouse and business intelligence metadata between warehouse tools, warehouse platforms and warehouse metadata repositories in distributed heterogeneous environment.

Data warehouse is a repository of an organization's electronically stored data. They are designed to manage and store the data and focuses on the usage of data to facilitate reporting and analysis. The purpose of a data warehouse is to house standardized, structured, consistent, integrated, correct, cleansed and timely data, extracted from various operational systems in an organization. The extracted data is integrated in the data warehouse environment in order to provide an enterprise wide perspective, one version of the truth. Data is structured in a way to specifically address the reporting and analytic requirements.

An essential component of a data warehouse system is the metadata and tools to manage and retrieve metadata.

Extensible Business reporting language (XBRL)

XBRL is a standards-based way to communicate and exchange business information between business systems. XBRL is XML-based. These communications are defined by metadata set out in XBRL taxonomies, which capture the definition of individual reporting concepts as well as the relationships between concepts and other semantic meaning.

Geographical Information Systems (ISO 19115)[15]

The Geographical Information Systems (ISO 19115 defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

It defines:

- mandatory and conditional metadata sections, metadata entities, and metadata elements;
- the minimum set of metadata required to serve the full range of metadata applications (data discovery, determining data fitness for use, data access, data transfer, and use of digital data);
- optional metadata elements, to allow for a more extensive standard description of geographic data, if required;
- a method for extending metadata to fit specialized needs.

It may be used for Cataloguing of datasets, clearinghouse activities, and the full description of datasets (geographic datasets, dataset series, and individual geographic features and feature properties). Though ISO 19115 is applicable to digital data, its principles can be extended to many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data.

General Data Dissemination System (IMF – GDDS)[16]

The General Data Dissemination System purposes are (i) to encourage member countries to improve data quality; (ii) to provide a framework for evaluating needs for data improvement and setting priorities in this respect; and (iii) to guide member countries in the provision to the public of comprehensive, timely, accessible, and reliable economic, financial, and socio-demographic statistics in a world of increasing economic and financial integration. The guidance comprises four dimensions:

- The data: coverage, periodicity, and timeliness
- Quality of the disseminated data
- Integrity of the disseminated data
- Access by the public

Special Data Dissemination Standard (IMF – SDDS)

The Special Data Dissemination Standard (SDDS) was established by the International Monetary Fund (IMF/Fund) to guide members that have, or that might seek, access to international capital markets in the provision of their economic and financial data to the public. Both the General Data Dissemination System (GDDS) and the SDDS are expected to enhance the availability of timely and comprehensive statistics and therefore contribute to the pursuit of sound macroeconomic policies; the SDDS is also expected to contribute to the improved functioning of financial markets.

---

[15] Source: http://www1.unece.org/stat/platform/display/metis/GIS+ISO+19115

[16] The general data dissemination system: http://dsbb.imf.org/images/pdfs/gdds_oct_2003.pdf and the general data dissemination system 'guide for participants and users' is available at: http://dsbb.imf.org/images/pdfs/gddsguide.pdf

The statistical standards applied in the UN-ESCWA region

This part is based on the questionnaire that UN-ESCWA member countries have filled in with regard to the statistical standards they apply or plan to use in their statistics work. UN-ESCWA identified 3 countries applying the following standards.

Table 1- Summary of the statistical standards applied in the UN-ESCWA region.

Metadata standards	Egypt	Jordan	Palestine	Syria
Dublin core (DC)	YES	YES	YES	NO
Data Documentation Initiative (DDI)	YES	YES	YES	NO
Metadata Registries (ISO 11179)	NO	YES	NO	NO
Statistical Data and Metadata Initiative (SDMX)	NO	YES	NO	YES
Common warehouse model ISO 19504	NO	YES	NO	NO
Extensible Business reporting language (XBRL)	NO	YES	NO	NO
Geographical Information Systems ISO 19115	YES		NO	YES

The following UN-ESCWA member countries were found to use the IMF's Data Quality Assessment Framework (DQAF)/ GDDS for metadata: Qatar[17], Lebanon[18], Bahrain[19], Iraq[20], Kuwait[21], Oman[22], Saudi Arabia[23], Sudan[24], Syria[25] and Yemen[26]. The following UN-ESCWA countries subscribed to the SDDS standard of the IMF's framework: Egypt[27], Jordan[28],

In addition Qatar uses the data warehouse system. Egypt uses the DDI based on the DCMI. Sudan reportss using Sudaninfo (devinfo), and Yemen reports using devinfo.

---

[17] Qatar dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=QAT
[18] Lebanon dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=LBN
[19] Bahrain dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=BHR
[20] Iraq dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=IRQ
[21] Kuwait dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=KWT
[22] Oman dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=OMN
[23] Saudi Arabia dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=SAU
[24] Sudan dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=SDN
[25] Syria dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=SYR
[26] Yemen dissemination DQAF at the IMF website http://dsbb.imf.org/Pages/GDDS/CtyCtgList.aspx?ctycode=YEM
[27] Egypt dissemination DQAF at IMF website http://dsbb.imf.org/Pages/SDDS/CtyCtgList.aspx?ctycode=EGY
[28] Jordan dissemination DQAF at IMF website http://dsbb.imf.org/Pages/SDDS/CtyCtgList.aspx?ctycode=JOR

Future plans in metadata in the UN-ESCWA region

Table 2- Summary of the statistical standards planned to be used in the UN-ESCWA region

Metadata standards	Egypt	Jordan	Palestine	Syria
Dublin core (DC)	YES	?	YES	-
Data Documentation Initiative (DDI)	YES		YES	-
Metadata Registries (ISO 11179)	NO		NO	-
Statistical Data and Metadata Initiative (SDMX – ISO 17369)	NO		YES	YES
Common warehouse model ISO 19504	NO		NO	NO
Extensible Business reporting language (XBRL)	NO		NO	-
Geographical Information Systems ISO 19115	YES		NO	YES

Table 3- Summary of the statistical standards known in the UN-ESCWA region

Metadata standards	Egypt	Jordan	Palestine	Syria
Dublin core (DC)	YES	YES	YES	NO
Data Documentation Initiative (DDI)	YES	YES	YES	NO
Metadata Registries (ISO 11179)	NO	YES	NO	NO
Statistical Data and Metadata Initiative (SDMX – ISO 17369)	NO	YES	YES	YES
Common warehouse model ISO 19504	NO	YES	NO	YES
Extensible Business reporting language (XBRL)	NO	YES	NO	NO
Geographical Information Systems ISO 19115	YES	-	NO	YES

Part III: The statistical business process and statistical metadata

Generic Statistical Business Process Model and Metadata management

There were two specific needs that led to the development of a unique Generic Statistical Business Process Model. This two linked together, and resulted in a model that covers several aspects of the statistical computing infrastructure.

- Introducing a common language, when discussing needs of statistical metadata and metadata information systems throughout a statistical survey cycle or census cycle. This aspect was tackled by the Joint UNECE / Eurostat / OECD Steering Group on Statistical Metadata (METIS)[29].
- Enabling exchange and sharing of statistical software tools and modules developed by national statistical offices. This aspect was tackled by the Joint UNECE / Eurostat / OECD Steering Group on Management of Statistical Information Systems (MSIS)[30].

The Generic Statistical Business Process Model finalized by the METIS Steering Group over the last few years, forms a part of a Common Metadata Framework (CMF). Part C of this framework is entitled "Metadata and the Statistical Cycle". This part refers to the phases of the statistical business process and provides generic terms to describe them. During a workshop to progress the development of Part C of the CMF, held in July 2007, the participants agreed that the model currently used by Statistics New Zealand, would provide a good basis for developing a "Generic Statistical Business Process Model" (GSBPM).. This current version of the model (version 4.0)[31], was approved by the METIS Steering Group for public release in April 2009.

The original intention was for the GSBPM to provide a basis for statistical organizations to agree on standard terminology to aid their discussions on developing statistical metadata systems and processes. The GSBPM should therefore be seen as a flexible tool to describe and define the set of business processes needed to produce official statistics.
 he Generic Statistical Business Process Model comprises 9 processing phases:

- Specify needs
- Design
- Build
- Collect
- Process
- Analyse
- Disseminate
- Archive
- Evaluate

---

[29] http://www.unece.org/stats/archive/04.01d.e.html
[30] http://www.unece.org/stats/archive/04.01a.e.html
[31] See: www.unece.org/stats/gsbpm

The GSBPM also recognizes several over-arching processes that apply throughout the nine phases, and across statistical business processes. Over-arching statistical processes include metadata management.

Metadata are generated and processed within each phase, there is, therefore, a strong requirement for a metadata management system to ensure that the appropriate metadata retain their links with data throughout the GSBPM;

Good metadata management is essential for the efficient operation of statistical business processes. Metadata are present in every phase, either created or carried forward from a previous phase. In the context of this model, the emphasis of the over-arching process of metadata management is on the creation and use of statistical metadata. The key challenge is to ensure that these metadata are captured as early as possible, and stored and transferred from phase to phase alongside the data they refer to. Metadata management strategy and systems are therefore vital to the operation of this model.

Metadata used at each phase of the GSBPM by NSOs of member countries

The following table summarizes the metadata used at each phase of the GSBPM as reported by the countries in their response to the UN-ESCWA questionnaire sent in 2011.

Phase		Metadata used		
		Palestine	**Egypt**	**Jordan**
Planning	1. Specify Needs	• Themes • Subject areas • Statistical object types • Datasets • Software • Access rights	• Take into account global variables • Set the required output • Set the software to be used • Access rights	• Global variables • Subject areas • Statistical objects types • Statistical Process description • Datasets • Server locations
	2. Design	• Object variables (identify the variables) • Classifications • Measuring units • data collection methodology • Access package	• Design output • design of work systems • classifications • context data elements • data collection methodology • concepts	• classifications • variables domains • measure units • context data elements • data collection methodology • software • data sets
	3. Build	• Question • Questionnaire • Data collection • Data collection Methodology	• Design and test of the questionnaire • Data collection Methodology	• Question • Questionnaire • Data collection Methodology…
Operation	4. Collect	• data collection methodology • questionnaire • collection strategy • statistical units	• Data collection methodology • questionnaire • data elements • classifications • statistical units • software • access rights	• Data collection Methodology • questionnaire • collection strategy • data elements • classifications • software…
	5. Process	• matrix	• Preparing the table	• statistical process

Phase		Metadata used		
		Palestine	**Egypt**	**Jordan**
		• register • statistical process • process implementation • operation implementation • derivation rules • computation implementation • classifications • data elements • statistical units	• coding and registration • data validation • design programs to extract tables • put probability weights • the preparation of the clean tape • derivation rules • process implementation • computation implementation • classifications	• process implementation • operation implementation • computation implementation • classifications • data elements…
	6. Analyse	• tables • statistical processes	• Prepare the preliminary report • validate outputs • use tables and graphs • statistical processes • software	• Tables • statistical processes • confidentiality rules • storage • software • access package ...
	7. Disseminate	• output releases • access rights • access packages • quality measures/ reports • systems and tools	• Updating dissemination tools • output releases • access rights • systems and tools of dissemination.	• analytical reports • output releases • quality measures/ reports • access rights • online catalogues ...
Post- processing	8. Archive	• Catalogues • server locations • access rights • access packages	• Archiving publications	• server locations • access rights • access packages ...
	9. Evaluate	• Guides • Budgets		• policies • guides • annual reports • budgets

Part IV: Metadata management software tools

This part will provide practical information for national statistical offices where to look for useful tools and experiences with focus on the experiences of Palestine and Egypt (Jordan, need to check with ESCWA) with the Microdata Management Toolkit.

Microdata management toolkit

1. Rationale, objectives and standards

The Microdata Management Toolkit[32] was developed by the World Bank Data Group for the International Household Survey Network (IHSN) and its main objective is to help data producers share their datasets with many users. The aim in developing the Toolkit is also to promote the adoption of standards for international microdata documentation, dissemination and preservation, as well as to foster best practices by data producers in developing countries. It complements other efforts by the International Household Survey Network (IHSN) to produce and distribute tools and guidelines for improved management and use of microdata.

The Toolkit aims to serve the needs of various users, including data producers, the research community, survey sponsors and data repositories. Data producers will benefit from using the Toolkit in a number of ways. First, they will gain from better data and metadata preservation. In addition, the Toolkit provides a repository for the institutional memory surrounding each data collection activity. The capacity of staff to document microdata is increased, which is also likely to help identify weaknesses in data collection and processing methods, and in turn improve future data collection. Finally, the Toolkit provides a tool for packaging and distributing micro-datasets.[33]

The Microdata Management Toolkit provides a user-friendly framework for microdata documentation. It uses two metadata specifications: the Data Documentation Initiative (DDI) and the Dublin Core Metadata Initiative (DC). Both are defined in the Extensible Markup Language (XML).

2. Structure and content of the toolkit

The Toolkit comprises two modules:

- The **Metadata Editor** is used to document and view data and metadata in a single Nesstar[34] format file in accordance with international standards. It is used to import data from various standard format (SPSS, ASCII, STATA…), and provide

---

[32] http://www.surveynetwork.org/home/index.php?q=tools/toolkit
[33] Microdata Management Toolkit, User's guide, IHSN
http://www.surveynetwork.org/toolkit/download/_file_storage/toolkit1.1/UsersGuide_EN.pdf
[34] http://www.nesstar.com/

comprehensive metadata in user-friendly screens, enabling users to view and re-export, all data and metadata to various formats, at once using the Nesstar explorer program to do so. The Nesstar program does not allow users to modify data and metadata.

- The **CD-ROM Builder** is used to generate user-friendly outputs that would contain not only the data and metadata, but also all external resources and possibly additional material for dissemination and archiving. It also produces a user-friendly HTML output for sharing and preserving data and metadata that can be distributed on CD-ROM, DVD, and Internet…

All components of the Toolkit are available free of charge to all users[35].

Figure 2 the metadata editor V1.2 core elements
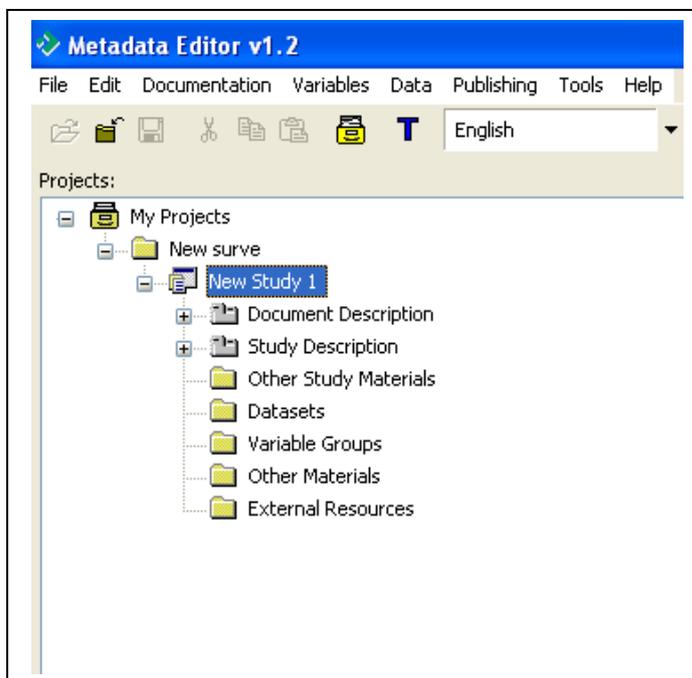
The main metadata editor elements are documentation description and study description.

---

[35] The Toolkit is distributed as a complete installation package on a CD-ROM. Alternatively; individual components can be retrieved and installed independently. The latest Information on how to obtain or download these products is available on the IHSN Toolkit web site at http://www.surveynetwork.org/toolkit.

Figure 3 the metadata editor V1.2 expended documentation description elements
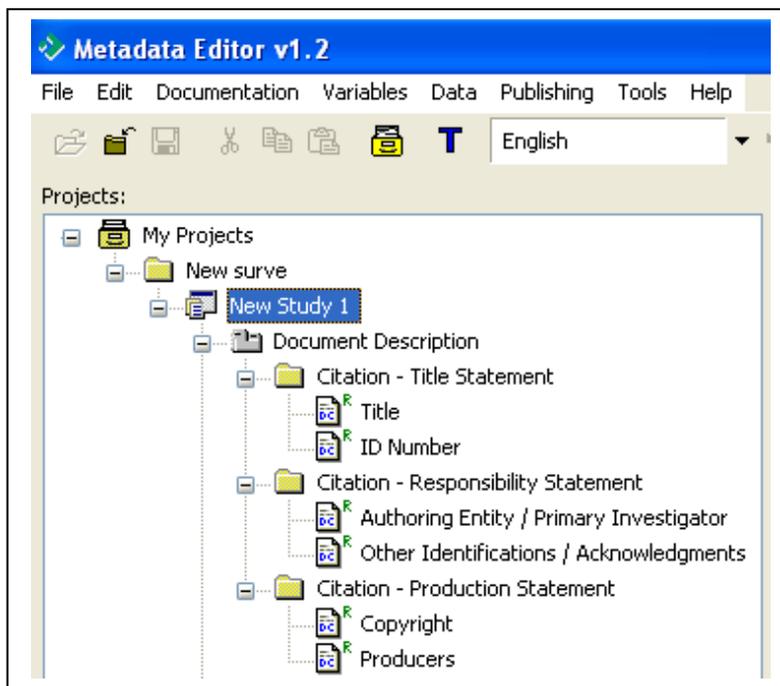
The expended metadata editor elements related to the document description are:

- Citation- Title statement
- Citation – responsibility statement
- Citation – production statement

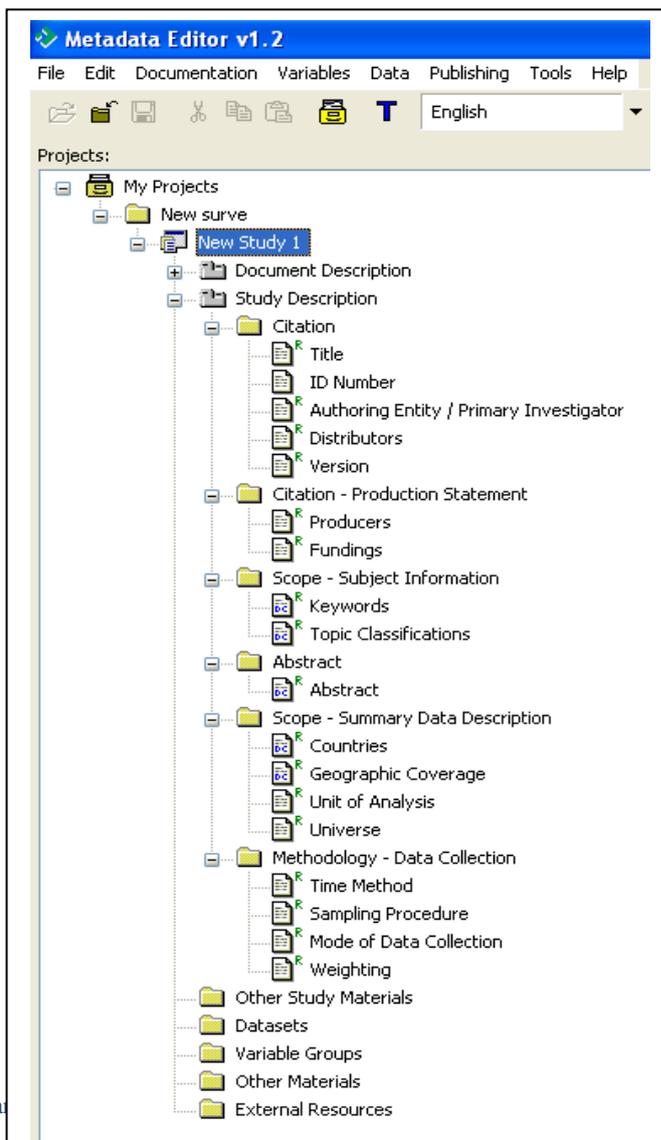Figure 4 the metadata editor V1.2 expended documentation description elements

The expended metadata editor elements related to the study description are:

- Citation- Title statement
- Citation – production statement
- Scope- Subject information
- Abstract
- Scope- Summary data description
- Methodology- Data collection
- Other study materials
- Datasets
- Variable Groups
- Other materials
- External resources

1. Using the toolkit

The Microdata Management Toolkit is designed to promote better use of microdata by improving their documentation, dissemination, and preservation.

3.1 Documentation

Documentation should accurately describe the data. The information should be clear so that the data are not incorrectly used. It should also be comprehensive, so that the statistical agency is not dependent on the institutional memory of staff. A basic principle is that all information that can foster the effective and accurate use of datasets by secondary users should be preserved and disseminated. Unfortunately, documentation is often the last step of the survey process, and it is then often too late to capture all metadata produced during the life cycle of the data collection activity. This result in the loss of useful information generated at early stages, such as the comments received from various stakeholders at the stage of questionnaire design, problems encountered during pilot-testing of the questionnaire, etc. Treating documentation as an ongoing part of survey activity will reduce the documentation costs and increase its quality.

The reason documentation is not routine is that it is typically difficult and time consuming to do well. Adoption of international metadata standards, such as the Data Documentation Initiative (DDI) and the Dublin Core (DC) specifications, can reduce the burden considerably, because they provide a rigorous framework for organizing the process. The Microdata Management Toolkit is a dedicated tool developed to allow easy implementation of the DDI and DC.

3.2 Dissemination

Providing secondary users with access to microdata can lead to many positive outcomes, benefiting both the research community and the data producer. These include:

- Increased value of the data through new and innovative research with greater diversity and scope.
- Improved methods of data collection and processing as a result of user feedback.
- Reduced duplication of data collection activities.
- Training junior researchers in the analysis of microdata.

Disseminating microdata also entails costs and risks. These include:

● Financial costs, although these are only a small fraction of the cost of collecting and processing data.
● Increased risk of a breach of confidentiality. This is a dominant issue from the point of view of official data producers, for whom maintaining the trust of respondents is crucial.
● Loss of data quality. Microdata dissemination may impose a reasonable level of anonymization of the data collected.
● Last, data producers are sometimes concerned that the quality of their microdata may not be good enough for public dissemination. Experience shows, however, that more criticism comes from the refusal to disseminate data than from the dissemination of less than perfect data. In most cases the benefits of dissemination outweigh the costs and risks, and official data producers in developing countries are increasingly adopting more open dissemination

policies. Proper protocols defining the procedures followed for archiving and disseminating microdata should be defined and adopted by all official data producers.

3.3 Preservation

Fully documenting and archiving datasets helps ensure that important survey data and metadata are preserved for future reference and analysis. This is particularly true in agencies with high rates of turnover.

The Microdata Management Toolkit preserves microdata and the related metadata in a common format. This standardization, which complies with the DDI standard, increases the likelihood that the information will be preserved. It also allows sharing of information with other DDI-compliant survey repositories. Finally, the Microdata Management Toolkit's CD-ROM Builder application provides an easy and inexpensive way to produce Internet and CD/DVD archives of data and metadata.

2. Using the Microdata Management Toolkit in the UN-ESCWA region

The case of Egypt

Since 2009, the Microdata Management Toolkit was used in a number of surveys and publications compiling administrative data sources and publications conducted by the CAPMAS or their partners. To date the CAPMAS has done this effort for the following ten surveys and statistical publications:

- Statistical indicators for the basic electronic measuring the information society, January 2009. (in Arabic: إحصاء المؤشرات الإلكترونية الأساسية لقياس م جتمع المعلومات شهر يناير 2009)
- Monthly Summary of Foreign Trade, December 2009. (in Arabic: الملخص الشهري للتجارة الخارجية لشهر ديسمبر 2009)
- Annual Bulletin of Statistics Industrial production in the private sector in 2008 (in Arabic: 2008 النشرة السنوية لإحصاء الإنتاج الصناعى فى منشآت القطاع الخاص لعام)
- Annual Bulletin of Statistics for the construction and building companies, public sector/ businessyear2008/2007 (in Arabic: النشرة السنوية لإحصاء التشييد والبناء لشركات القطاع العام /الأعمال العام 2008/2007)
- Annual Bulletin of Statistics education in schools, colleges and training centers are not subject to the Ministries of Education, Al-Azhar 2007/2008 (in Arabic: النشرة السنوية لإحصاءات التعليم بالمدارس والمعاهد ومراكز التدريب غير الخاضعة لوزارات التعليم والأزهر 2007/2008)
- Annual Bulletin of marriage and divorces statistics in 2008 (in Arabic: النشرة السنوية لإحصاءات الزواج و الطلاق عام 2008)
- Annual publication of statistics of births and deaths in 2008 (in Arabic: النشرة السنوية لإحصاءات الموالي والوفيات عام 2008)
- Annual Bulletin of Statistics and financial indicators for public sector companies and public sector(excluding banks and insurance companies2009/2008) (in Arabic: النشرة السنوية للإحصاءات والمؤشرات المالية لشركات قطاع الأعمال العام وال قطاع العام (عدا البنوك وشركات التأمين2009/2008))

- Sample survey of employment, annual consolidated results for 2007 (in Arabic: بحث العمالة بالعينة- 2007, النتائج السنويه المجمعه لعام 2007)
- The labor force survey in 2009,annual publication for the year 2009 (in Arabic: بحث القوى العاملة لعام 2009, النشرة السنوية المجمعة لبحث القوى العاملة لعام 2009)

In general each statistical product (survey, administrative data, and statistical publication) published under the CAPMAS metadata specific tab, has the basic Microdata Management Toolkit requirements or tabs. The table X below presents in details each statistical product and Microdata Management Toolkit requirements and elements

Table x Assessing the use of the Microdata management toolkit in Egypt (as of Dec 1, 2011)

Survey	Overview	Technical information:						Data sets		Access policy	Search by indicator	Tables & reports		Reports
		Sample	(Survey) questionnaires	Data collection	Data preparation	Data evaluation	Technical documentation	Data file	Indicators			Statistical tables	Reports	
Statistical indicators for the basic electronic measuring the information society, January 2009	Detailed	Yes	Yes	Yes	Yes	Yes	Yes as Pdf file	Variable descriptions & figures for selected number of variables	Yes as Pdf file	Yes	Yes	Yes	No	Survey background, concept definitions & detailed tables as Pdf file
Monthly Summary of Foreign Trade, December 2009	Detailed	No	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Annual Bulletin of Statistics Industrial production in the private sector in 2008	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annual Bulletin of Statistics for the construction and building companies, public sector/ business year2008/2007	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annual Bulletin of Statistics education in schools, colleges and training centers are not subject to the Ministries of Education, Al-Azhar 2007/2008	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annual Bulletin of marriage and divorces statistics in 2008	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annual publication of statistics of births and deaths in 2008	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annual Bulletin of Statistics and financial indicators for public sector companies and public sector(excluding banks and insurance companies2009/2008)	Yes	No	Yes	Yes	No	No	No	?	?	Yes	Yes	Yes	Yes	Yes
Sample survey of employment, annual consolidated results for 2007														
The labor force survey in 2009,annual publication for the year 2009														

Figure 5- example of the Microdata Management Toolkit used at CAPMAS for the Statistical indicators for the basic electronic measuring the information society, January 2009

The case of Palestine

The Palestinian Central Bureau of Statistics (PCBS) used the Microdata Management Toolkit in 36 surveys and publications compiling administrative data sources and publications conducted by the PCBS or their partners:

- Construction Contractors Survey, Twelfth Volume
- Construction Contractors Survey–2008, fifteenth Round
- Domestic Outbound Tourism Survey 2009
- Environmental Economic Survey 2009, Fifth Round
- Environmental Economic Survey, 2006
- Environmental Economic Survey, 2008
- Environmental Survey for Education Sector, 2008
- Environmental Survey for Education Sector, 2010
- Environmental Survey for Health Care Centers 2009
- Environmental Survey for Health Care, 2007
- Established Municipal Data Base in Palestinian Territory, 2009
- Expenditure and Consumption Survey 2007
- Expenditure and Consumption Survey 2009
- Household Cultural Survey 2009
- household Energy Survey, July 2010
- Household Energy Survey 2008
- Housing Conditions Survey 2006
- Industrial Survey 2008
- Internal Trade Survey 2005
- Internal Trade Survey 2006
- Labor Force Survey 2008
- Olive Presses Survey 2005, eleventh issue
- Olive Presses Survey 2006, twelfth volume
- Olive Presses Survey 2007
- Olive Presses Survey 2008
- Population, Housing and Establishment Census 2007, Census 2007
- Services Survey 2005
- Services Survey 2006
- Services Survey 2008
- The Establishment Census 2007
- The Industrial Survey – 2005, twelfth volume
- Transport Storage and Communication Survey 2005
- Transport Survey Informal Sector 2005
- Transport Survey: Informal Sector 2006
- Transport, Storage and Communication Survey 2006
- Victimization Survey, 2008

In general each statistical product (survey, administrative data, and statistical publication) published under the PCBS metadata specific tab, has the basic Microdata Management Toolkit requirements or tabs. In addition the number of the version, each statistical product had the following items:

Overview

Technical information:

- Sampling
- (Survey) questionnaires
- Data collection
- Data processing
- Data appraisal
- Technical documentation

Data sets

- Data file
- Variable groups

Access policy

Variable search

Other materials

Tables and reports

- Statistical tables
- Reports

Reports

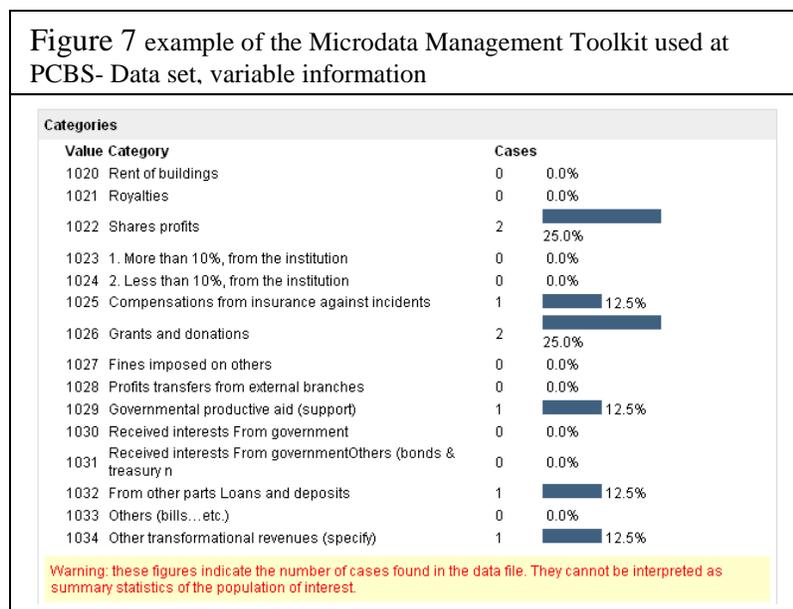Figure 6 - example of the Microdata Management Toolkit used at PCBS
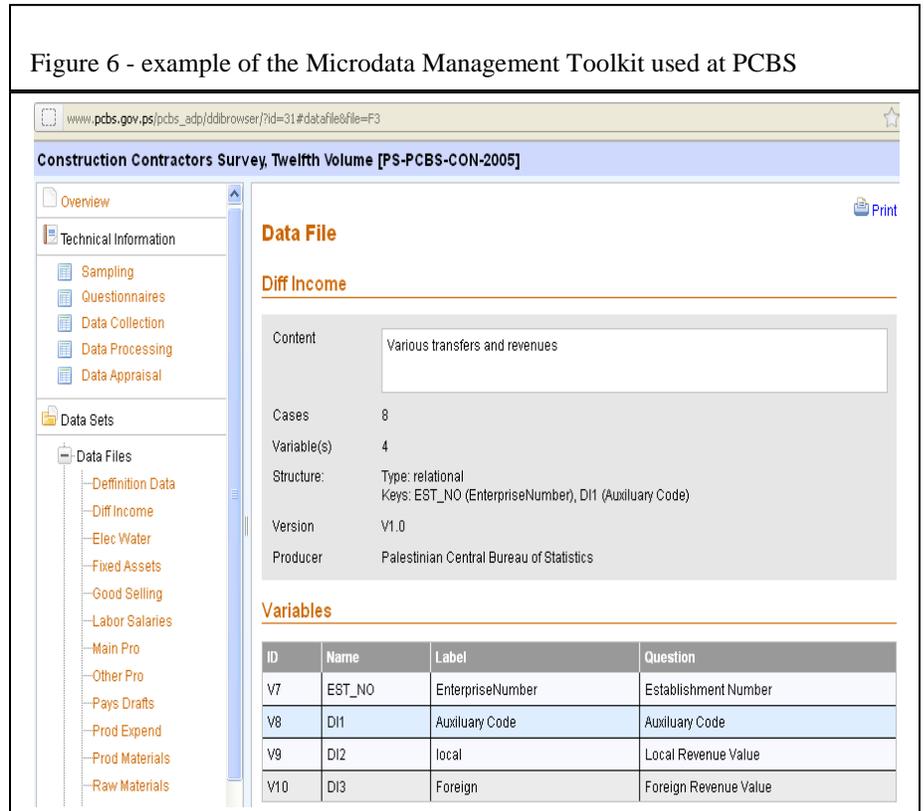
Figure 7 example of the Microdata Management Toolkit used at PCBS- Data set, variable information

In addition to the Microdata Management Toolkit, the PCBS is also using the DevInfo as a tool for organizing, storing and presenting the 2010 Census data[36] and the MDG indicators[37].

---

[36] http://www.pcbs.gov.ps/CensusInfo/
[37] http://www.pcbs.gov.ps/Palinfo/home.aspx

Conclusion

This handbook provided a review of the current situation of the UN-ESCWA member states in the area of metadata.

It is obvious that the situation is different from one country to another is different. Few NSOs have been investing in Metadata development while other NSO's have not started yet or at least have not disseminated the information outside the office. There may be some countries who started working on the subject but have not considered it as a priority to the office.

A lot of work and investment is highly needed in this field. This need have clearly been reflected in the recommendations that came out of the Expert Group Meeting on Data and Metadata Reporting that was organized by the UN-ESCWA in Amman, Jordan, 16-18 July 2011.

These recommendations for member countries were:

1.   The participants stress a need for Arab National Statistical Offices (NSOs) to have a legal framework that would institutionalise documenting statistical data, processes and surveys.

2.   The participants encourage all NSOs to create teams for documenting statistical data, processes and surveys, and that such documentation is stored in central repositories.

3.   The participants call upon NSOs and international organizations to standardise classifications and concepts used in metadata within the Arab region.

4.   The participants urge NSOs to follow international and national standards as a prerequisite for quality assurance in statistics.

5.   The participants encourage NSOs that positive experience of NSOs of Egypt, Jordan and Palestine in tools (Toolkit) and processes for metadata management are brought to reality in other NSOs.

These recommendations for the ESCWA Secretariat were:

6.   The participants asked that ESCWA creates a registry of questionnaires periodically issued by international organizations (using the current knowledge about metadata registries).

7.   The participants asked ESCWA to speed up implementation of SDMX in the region.

8.   The participants asked that ESCWA Secretariat supports capacity building in statistical metadata through:
     - creating a forum for exchange of experience and expertise;
     - Organizing training and practical workshops;
     - Organizing e-learning.

9. The participants asked ESCWA to organize Trainings of Trainers (ToT) on standards and tools for data and metadata exchange.

10. The participants asked ESCWA to undertake activities aimed at increasing awareness about standards and tools.

11. The participants, considering that the Microdata Management Toolkit has a potential for wide utilisation across the region for documenting at micro level, asked that ESCWA includes the Toolkit in its capacity development programmes.

12. The participants asked that ESCWA works with member countries to map DDI with SDMX making it possible to utilise both systems as appropriate.

References and useful links

- Metadata standards at the International Household Survey Network (IHSN) http://www.internationalsurveynetwork.org/home/index.php?q=tools/documentation/standards

- The metadata toolkit manual is available at the following: http://www.surveynetwork.org/toolkit/download/_file_storage/toolkit1.1/UsersGuide_EN.pdf

- The General Data Dissemination System 'guide for participants and users' (2007), The International Monetary Fund (IMF): http://dsbb.imf.org/images/pdfs/gddsguide.pdf

- Framework for the specification and standardization of data elements (1999), ISO and IEC http://metadata-stds.org/11179-1/ISO-IEC_11179-1_1999_IS_E.pdf

- The Nesstar webpage: http://www.nesstar.com/

- Common Metadata Framework Part A: Statistical Metadata in a Corporate Context: http://www1.unece.org/stat/platform/display/metis/Part+A+-+Statistical+Metadata+in+a+Corporate+Context

- Metadata standards, concepts, models, best practices and other methodological materials Part B: http://www1.unece.org/stat/platform/display/metis/Part+B+-+Metadata+Concepts%2C+Standards%2C+Models+and+Registries

- Metadata and statistical business process Part C: http://www1.unece.org/stat/platform/display/metis/Part+C+-+Metadata+and+the+Statistical+Business+Process

- Metadata implementations and cases studies: http://www1.unece.org/stat/platform/display/metis/Part+D+-+Implementation

- The Generic statistical business process model: http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model

- Statistical Data and Metadata eXchange initiative: http://www.sdmx.org

Annexes

1. UN-ESCWA member countries NSO's websites

Member state	NSO official name	NSO website
Egypt	Central Agency for Public Mobilization and Statistics (CAPMAS)	http://www.capmas.gov.eg/
Bahrain	Bahrain Central Informatics Organization	http://www.cio.gov.bh/CIO_ENG/default.aspx
Iraq	Central Organization for Statistics and Information Technology (COSIT)	http://cosit.gov.iq/english/index.php
Jordan	Ministry of Planning, Department of Statistics	http://www.dos.gov.jo/dos_home_a/main/index.htm
Kuwait	Central Statistical Organization	http://www.cso.gov.kw/
Lebanon	Central Administration of Statistics	http://www.cas.gov.lb/
Oman	Ministry of National Economy	http://www.mone.gov.om/Default-ar.aspx
Palestine	Palestinian Central Bureau of Statistics	http://www.pcbs.gov.ps/
Qatar	Statistical Authority	http://www.qsa.gov.qa/eng/index.htm
Saudi Arabia	Central Department of Statistics	http://www.cdsi.gov.sa/
Sudan	Central Bureau of Statistics	http://www.cbs.gov.sd/
Syria	Central Bureau of Statistics	http://www.cbssyr.org/
United Arab Emirates	Central Directorate of Statistics	http://www.economy.gov.ae/Arabic/Pages/default.aspx
Yemen	Central Statistical Organization	http://cso-yemen.org/index.php?lng=english&/

2. Core principles for metadata management

Part A of the Common Metadata Framework[38] identifies the following sixteen core principles for metadata management, all of which are intended to be covered in the over-arching Metadata Management process, and taken into the consideration when preparing the statistical metadata system (SMS) vision and global architecture, and when implementing the SMS. The principles can be presented in the following groups:

Metadata handling	i.	***Statistical Business Process Model***: Manage metadata with a focus on the overall statistical business process model.
	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
	iv.	***Versions***: Preserve history (old versions) of metadata.
Metadata Authority	i.	***Registration***: 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.
	ii.	***Single source***: Ensure that a single, authoritative source ('registration authority') for each metadata element exists.
	iii.	***One entry/update***: Minimize errors by entering once and updating in one place.
	iv.	***Standards variations***: Ensure that variations from standards are tightly managed/approved, documented and visible.
Relationship to Statistical Cycle / Processes	i.	***Integrity***: Make metadata-related work an integral part of business processes across the organization.
	ii.	***Matching metadata***: Ensure that metadata presented to the end-users match the metadata that drove the business process or were created during the process.
	iii.	***Describe flow***: Describe metadata flow with the statistical and business processes (alongside the data flow and business logic).
	iv.	***Capture at source***: Capture metadata at their source, preferably automatically as a bi-product of other processes.
	v.	***Exchange and use***: 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

---

[38] See: http://www.unece.org/stats/cmf/PartA.html

as XML.

Users	**i.**	***Identify users***: Ensure that users are clearly identified for all metadata processes, and that all metadata capturing will create value for them.
	ii.	***Different formats***: The diversity of metadata is recognized 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.
	iii.	***Availability***: Ensure that metadata are readily available and useable in the context of the users' information needs (whether an internal or external user).