Documentation Metatadata of household survey Based on DII and DCMI standards
(the experience of CAPMAS in this field)
Waleed Ameen Abd Elkhalik Mohamed
CAPMAS, Demographics Statistics
Cairo, Egypt.
Email: woo2005xp@yahoo.com or kant2012xp@hotmail.com

Abstract:

Metadata is "data about data". Metadata provides documentation about data elements or attributes, (name, size, data type, etc), about records or data structures (length, fields, columns, etc), and about data itself (where it is located, how it is associated, ownership) including descriptive information about the context, quality and condition, or characteristics of the data , instructions for how data were collected, definitions for data items. on the other hand Metadata is essential when data are to be used by those not familiar with the sources, methods, and details of a database that are necessary to fully understand appropriate use and interpretation of findings. Duval, E. (2001), This paper will illustrate What is metadata and why is it important?, and will describe Standards of documentation the statistical survey : (DCMI)" Dublin core metadata initiative" – (DDI) "data documentation initiative" from the prospective of CAPMAS (the National office for statistics at Egypt) experience at using toolkit program in documentation and how Documentation metadata solve many problem at the statistical system such as availability of row data , quality control , Building institutional memory by documentation, dissemination and preservation of statistical survey, In this context ,we can say that documentation Metadata of statistical survey is key to ensuring that statistical resources will survive and continue to be accessible into the future because as well as to foster best practices by data producers in developing countries.

Key Words: *Micro Data; Metadata Editor; Standards of Documentation; Toolkit.*

1. Introduction:

Metadata is often defined as data about data. It is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource", especially in a distributed network environment like for example the internet or an organization. A good example of metadata is the cataloging system found in libraries, which records for example the author, title, subject, and location on the shelf of a resource. Statistical metadata is structured information about statistics. This includes information used for producing, disseminating, understanding, finding and (re)using statistics. National Information Standards Organization (2007),

2- Using Toolkit in Documentation: (Toolkit user's guide (2006)).

The Micro data Management Toolkit developed by the World Bank Data Group for the International Household Survey Network (IHSN) is designed to documentation metadata for statistical survey, so the following we highlight on how to use this program at documentation,

2.1 Components of micro data management toolkit:

Toolkit has three sections:

- Metadata Editor

  We can use this tool to Document Survey Data In Accordance International Standards.

- CD-Rom builder

  We can use this tool to generate user-friendly outputs, such as CDS, websites, for dissemination and archiving

- The Explorer : We can use this tool:
  - For viewing metadata
  - For re-exporting data to various formats

2.2 Benefits of using toolkit : (IHFAN Quick Reference Guide (2008))

We can summarize the benefits of using toolkit in the following point:
- User friendly software for micro data.

- Facilitate metadata exchange (DDI, Dublin Core).
- Facilitate archiving (metadata and data, quality control).
- Facilitate preservation/dissemination: network, CD / DVD, Web Sites.
- Works with common data formats .
- Free or inexpensive .
- Availability of technical support and training.
- Supported by national, international and research communities .

On the other hand data producer also can get the following advantage from using toolkit in documentation:

- They will gain from better data and metadata preservation. in addition, the toolkit provides them with the institutional memory surrounding each data collection activity.
- It help them to identify weaknesses in data collection and processing methods.
- Toolkit also can provided them with a tool for packaging and distributing micro-datasets for all user.

3- Standards of Documentation (DDI) & (DCMI)

Documentation using toolkit is organized according to two standards (DDI&DCMI) each one of them includes many of the elements.

3.1 The first standard is DDI Data Documentation Intuitive

DDI Elements are organized in five sections:

3.1.1- Document Description

This section includes information about study title – metadata producer- data of production- DDI documentation version- DDI Document ID number.

3.1.2- STUDY DESCRIPTION

This section includes information about how the study should be cited, who collected, compiled and distributes the data, a summary (abstract) of the content of the data, information on data collection methods and processing.

3.1.3 DATA FILE DESCRIPTION

This section is used to describe each data file in terms of content, record and variable counts, version , producer.

3.1.4 VARIABLE DESCRIPTION

This section presents detailed information on each variable, including literal question text, universe, variable and value labels, derivation and imputation methods, and soon.

3.1.5- OTHER MATERIAL

This section allows for the description of other materials related to the study. It has many elements such as documents (questionnaires, coding information, technical and analytical reports, interviewer's manuals, and so on).(**IHFAN Quick Reference Guide for Health Facility Assessment Data Archivists(2008)).**

3.2- Second Standard is DCMI (Dublin Core Metadata Initiative)

Dublin Core metadata standard is based on the same principles as the DDI specification. it organized to form an xml file.

Dublin Core Metadata Initiative (DCMI)

3.2.1- Title. the name by which the resource is formally known.

3.2.2- Subject. the topic of the resource.

3.2.3- Description. An abstract, a table of contents.

3.2.4- Type. the nature of the content of the resource (a survey questionnaire, a data processing syntax program, a map).

3.2.5- Relation. a reference to a related resource

3.2.6- Coverage. the extent or scope of the content of the resource.

3.2.7- Creator. the person(s), organization(s), or service(s) responsible for making the content of the resource

3.2.8- Publisher. the person(s), organization(s), or service(s) responsible for making the resource available.

3.2.9-Contributor. The person(s), organization(s), or service(s) having contributed to the content of the resource.

3.2.10- Rights. a rights management statement for the resource.

3.2.11- Language. a language of the intellectual content of the resource. (Available at: http://dublincore.org/)

4- Gathering and preparing the study documentation

All information related to the study may be useful and should be archived (even if not all will be disseminated to the public). This includes not only technical documents such as the questionnaires or list of codes (obviously needed by data users), but also administrative reports (potentially useful for implementation of future surveys), and other documents such as a compilation of the comments provided by stakeholders at the time the questionnaire was designed, etc.

We can summarize the Resources that be included in Documentation as following:

- The study questionnaire(s); make sure that the cover page and all sections are included. If the questionnaire exists in multiple languages, provide all versions.
- All technical, analytical and administrative documents such as the followings:
- Sampling information; *-Interviewers and supervisors manuals;
- List of codes;   *-Instructions for data editing;
- Study report;   *- Tabulation and analysis plans;
- Analytical papers and policy briefs that made use of the data
- Survey budget and other key planning documents;
- PowerPoint presentations and other related material;
- Computer programs (used for data entry, editing, tabulation and analysis);
- Photos;   *-Tables- Maps flyers, videos, posters, songs, etc.).

5- Importance of documentation

- Improved methods of data collection and processing as a result of user feedback
- Training junior researchers in the analysis of microdata.
- Facilitate quality control and reporting by implementing agencies
- Help us to identify weaknesses in data collection and processing methods, and in turn improve future data collection.
- distributing micro-datasets.

On the other hand Documentation, or metadata, helps the researcher to

- Find the data they are interested in. Without names, abstracts, keywords and other important metadata element it might be difficult for a researcher to locate the datasets and variables that meet his or her research requirements.
- Understand what the data are measuring and how the data have been created. Without proper descriptions of the design of the survey and the methods used

when collecting and processing the data, the risk is high that the user will misunderstand and even misuse them.

☐ Assess the quality of the data. Information about the data collection standards, as well as any deviations from the planned standards, is important knowledge for any researcher who wants to know whether the data are useful for his project.

6- Documentation statistical survey based on DDI&DCMI standards and experience of CAPMAS(National statistical office in Egypt) in this field

Microdata Management Toolkit developed by the World Bank Data Group for the International Household Survey Network (IHSN) is designed to address the technical issues facing data producers. The aim in developing the Toolkit is 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 IHSN to produce and distribute tools and guidelines for improved management and use of Microdata.

In this context Documentation started at CAPMAS in Dcember 2009 when OCED send international expert to train 20 employee on documentation ,archiving ,dissemination four statistical survey (Income and expenditure, education, electronic indicators, birth and death) .

After that CAPMAS has been started archiving statistical survey with determine definations of Documentaion, why we need documentation of statistical survey? How we make documentation? Beacause this questions are very important to determine steps towards agood documentation for statistical survey :

1- Documentation : All information related with statistical survey .

2- Why documentation ? it help us to return for any material concerning with.survey.

3- How documentation? there are procedures must person who documentation flow it.

4- When decumentation? After completion statistical survey.

The following figure showing steps of using toolkit in documentation : (Micro data management Toolkit user's guide (2006)).
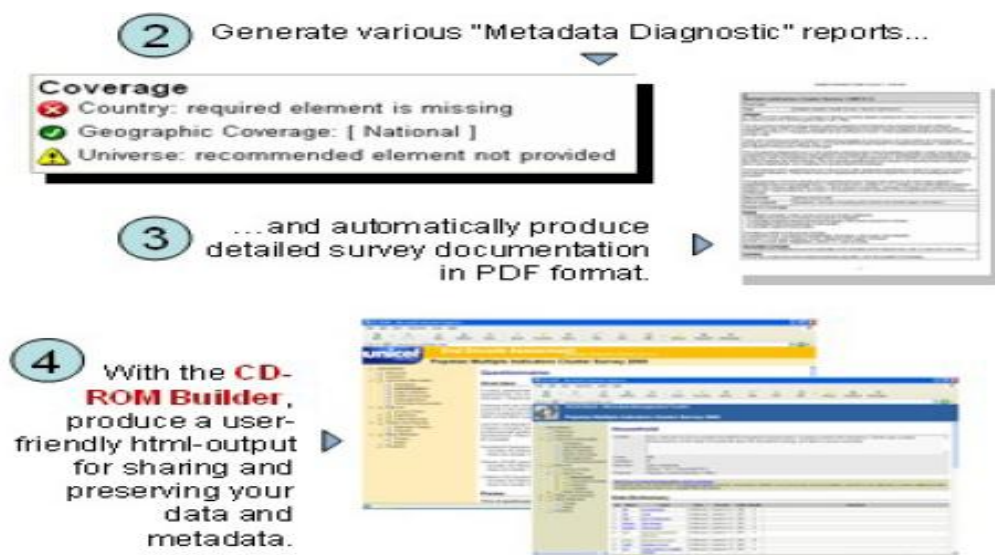
Shape (1) **steps of using toolkit in documentation**

Table (1):Total statistics that documented (published internal / external Dissemination) at CAPMAS 2010

	Name of the statistical survey	Dissemination of an internal	Dissemination of external
1	Labor force survey	2008-2010	2007-2009
2	Statistics of marriage and divorce.		2008-2008
3	statistics of births and deaths..	2007-2009	2008
4	Annual Bulletin of Statistics Industrial production in the private sector.	2007-2009	2008
5	Statistics of industrial production facilities	2008-2009	—
6	Statistics form the basic electronic indicators to measure the information society, the family.	2007-2008-2010	2009
7	Statistics of education	2006-2007	2007-2008
8	Count activity hotel and tourist villages in the sectors of public and private	2008-2009	—
9	Statistics of the building and construction companies to the public sector / business	2008-2009	2007-2008
10	Foreign Trade data	2008-2010	2009
11	financial indicators for public sector companies		2008-2009
12	Statistics and financial indicators for the organized private sector companies.	2009	⸺
13	Income and expenditure and consumption.	2008-2009	___
14	Survey of Health Services, 2009.	2009	___
15	survey of the crop area and production plant	2009	_____
16	Bulletin of public transport of passengers	2009	___
	Total = 33 survey		___

Conclusion

Toolkit help us to Gain better data and metadata preservation. In addition, the Toolkit provides National statistical offices with a repository for the institutional memory surrounding each data collection activity. The capacity of staff to document micro data 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.

Recommendations

There are some lessons we learned it from documentation using toolkit such as :

1- We must spread awareness with the importance of documentation.

2- We can used toolkit in development the scientific skills.

3- We must start documentation when we start collecting data from stakeholders.

REFERENCES

1- DDI and SDMX(2007): Complementary, Not Competing, Standards , Version 1.0, July 2007.

2- Dublin Core Metadata Initiative (DCMI): **Available at:** http://dublincore.org/**)** . (Accessed April 2013.)

3- Duval, E. (2001), Metadata Standards, What, Who & Why, Journal of Universal Computer Science, 7(7), 591-601.

4- IHFAN Quick Reference Guide for Health Facility Assessment Data Archivists(2008), DRAFT - Version.02

5- Micro data management Toolkit user's guide (2006), version 1.0 .Available in PDF format at www.surveynetwork.org/toolkit. (Accessed December 2010.)

6- National Information Standards Organization (NISO) (2007), A Framework of Guidance for Building Good Digital Collections, 3rd ed., A NISO Recommended Practice, NISO, Baltimore, USA