STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE	COMMISSION OF THE EUROPEAN COMMUNITIES
CONFERENCE OF EUROPEAN STATISTICIANS	EUROSTAT

Joint UNECE/Eurostat Work Session on Statistical Metadata (6 - 8 March 2002, Luxembourg)	Working Paper No. 21 English only

Topic (iii): Metadata and quality

USE OF METADATA IN DATA QUALITY ASSESSMENTS – AN SDDS EXAMPLE

Submitted by the International Monetary Fund (IMF)[1]

Invited paper

I.   INTRODUCTION

1.   In the last decade, increasing emphasis on transparency in the compilation and dissemination of statistics has provided the impetus for a proliferation of statistical metadata in various forms. Responding to the international community's call for standards and codes, the International Monetary Fund (IMF), as part of its two-tier Data Standards Initiative, established the Special Data Dissemination Standard (SDDS) in 1996 to guide its member countries that have, or that might seek, access to international capital markets in the dissemination of economic and financial data to the public.[2] Drawing on international "best practices", the SDDS provides a framework for the dissemination of information on statistical practices (metadata) built around four dimensions of data dissemination: data (i.e., coverage, periodicity, and timeliness); access by the public; the integrity of the data; and the quality of the data. SDDS metadata are posted on the Dissemination Standards Bulletin Board (DSBB), which is linked to actual country data (National Summary Data Pages) disseminated by SDDS subscribers on the Internet. These links facilitate monitoring of observance of the standard by the IMF, and they provide data access to financial markets and other data users. Currently, there are fifty subscribers to the SDDS and their metadata are posted on the DSBB. Also, the European Central Bank (ECB) and Eurostat have adopted the SDDS format for their metadata presentations on the Internet.

2.   The object of this paper is two-fold. First, it demonstrates that coupling a comprehensive and readily recognizable catalogue metadata framework with an assessment tool such as the Data Quality Assessment Framework (DQAF) provides the necessary and sufficient conditions for the conduct of data quality assessments benchmarked against international best practices. Second, it explains how an open exchange system for the dissemination of statistical information on the Internet could provide the platform for intelligent search capabilities in a pre-defined format using the DQAF.

---

[1] Prepared by Robert Di Calogero, Gary Jones, Paul Austin, Michel Le Marois.

[2] In 1997, the IMF launched the second tier of its Data Standards Initiative- the General Data Dissemination System (GDDS) to guide all other member countries in the provision of data to the public. The dimensions of the GDDS are similar to those of the SDDS. Thus, while this paper focuses on the SDDS metadata format, it is also applicable to the GDDS.

II.   THE SDDS METADATA FRAMEWORK AND DATA QUALITY ASSESSMENTS

3.   Since the launching of the DSBB in September 1996, subscription to the SDDS has grown to 50 countries, and the IMF is working with a number of other countries towards subscription, some of whom are already utilizing the SDDS format as the basis for disseminating statistical metadata. Additionally, the European Central Bank and Eurostat have adopted the SDDS format for their metadata presentations, and have provided links to the DSBB.[3] Further, metadata for 39 participants in the IMF's General Data Dissemination System (GDDS) built on the similar four dimensions—data characteristics, quality, access, and integrity—are posted on the DSBB, and the IMF is working with a number of additional countries that have committed to using the GDDS as a framework for the development of their national statistical systems. Together, this means that almost half of the IMF's 183 members publicly disseminate information about their statistical practices on the DSBB using the SDDS/GDDS framework.

4.   As a dissemination standard, the SDDS focuses on the disclosure elements of statistical best practices, and was crafted in response to the lessons of the international financial crisis of the mid-1990s, which shaped the IMF's early work on standards and codes. In recent years, a number of factors, including financial crises in emerging markets and the growth of Internet-based modes of data and metadata dissemination, have led to widespread recognition of the need for precise and robust tools to promote information access and data transparency. Such tools would leverage the potential analytical benefits of vast stores of unmined statistical information available on the Internet to assess data quality. In this context, the IMF's Statistics Department – encouraged by the experiences gained in the implementation of the SDDS and the GDDS– recognized the need not only to complement and extend the quality dimension of the SDDS and the GDDS, but more broadly to develop an objective, structured framework- based on international concepts and methodologies- for assessing the quality of statistical systems and outputs.

5.   The SDDS provides a catalogue metadata platform that addresses data quality issues in a structured, self-disclosure format. For several years, work on an *assessment* tool that would provide structure and a common language for the assessment of data quality has been underway in the IMF's Statistics Department, in consultation with national statistical offices, international organizations, and data users outside the Fund. This work has culminated in the development of the Data Quality Assessment Framework (DQAF).

6.   The DQAF brings together best practices and internationally accepted concepts and definitions in statistics to facilitate assessments of national practice in five dimensions of data quality, i.e., integrity, methodological soundness, accuracy and reliability, serviceability, and accessibility, as well as the related institutional prerequisites of quality (i.e., an analysis of the legal and institutional environment). In rendering metadata amenable to comparison with international best practices, the DQAF represents an evolution of the approach to data quality from that of *monitorable proxies* developed for the SDDS in 1996, to *observable features* that encompass all aspects of statistical systems and outputs.

7.   The DQAF is intended as a tool to assess the quality of the collection, production, and dissemination of data. At the highest level, the four dimensions of the SDDS consider the same characteristics as the five dimensions of the DQAF. However, reflecting the differing purposes of the SDDS and DQAF, the nature of the elements in the SDDS and the DQAF differ.[4] First, the SDDS typically *prescribes* a set of practices, while the DQAF *identifies* practices that contribute to the quality of data and are, therefore, relevant in assessing data quality. Second, the SDDS focuses on practices related to dissemination, while the DQAF includes, in addition, practices related to the collection and production of data. Each of the five quality

---

[3] The ECB's Euro metadata are accessible at http://dsbbimf.org/euronote.htm. Eurostat's Euro indicators are accessible at http://dsbb.imf.org/euroindicators.htm.
[4] Appendix 1 further illustrates the differences in purpose and scope in each of the SDDS/DQAF dimensions.

dimensions in the DQAF is broken down into a number of elements, and each element is assigned a number of indicators or pointers designed to draw out good practices.

8.   Data quality assessment tools, such as the DQAF, require a substantial amount of detailed metadata, beyond those available in the SDDS framework. A preliminary analysis of the SDDS metadata in terms of indicators vis-à-vis the DQAF dimensions (see Table 1) shows the extent of coverage in the SDDS, as follows:

- The "Methodological soundness" is the DQAF dimension for which the SDDS metadata provide the most extensive coverage. The "coverage characteristics" cell and the summary methodology statements provide comprehensive information on the application of international standards, guidelines, and agreed practices.

- The "Accessibility" and "Serviceability" DQAF dimensions are also adequately covered by the SDDS metadata. The SDDS framework, as a dissemination standard, provides detailed metadata on the availability of information to users; and the extent to which data are relevant, produced and disseminated in a timely fashion with appropriate periodicity, and follow a predictable revisions policy.

- The "Accuracy and reliability" DQAF dimension is not as adequately covered by the SDDS metadata, as the SDDS framework is not designed to provide detailed technical information on the quality of the source data, statistical techniques, and supporting assessments and validation.

- The "Prerequisites" and " Integrity" are the DQAF dimensions for which the SDDS metadata provides the least coverage. The SDDS appears less equipped to provide detailed metadata on conditions within the agency in charge of producing statistics that have an impact on data quality (such as the legal and institutional environment, resources, and quality awareness), as well as professionalism and ethical standards.

9.   In applying the DQAF in the context of data modules of the Report on the Observance of Standards and Codes (ROSC),[5] the IMF utilizes metadata (from the SDDS or the GDDS) where available, supplemented with information made available through discussions with the authorities, surveys of data users, the Internet and other publicly available information. Given the increasing importance of the Internet in disseminating statistical information, spurred on by the success of the DSBB, the capacity to access and repurpose this information to inform data quality assessments is imperative. The wider applications of the DQAF assessment methodology, including its usage for the derivation of Statistical Capacity Building Indicators by the PARIS21 Task Team on Statistical Capacity Building Indicators, underline the potential benefits of automated interrogation and content aggregation of statistical metadata systems for the statistical community. The development of standards and protocols for the efficient sharing and dissemination of statistical information on the Internet is necessary for realizing such potential, and these efforts are already being pursued under a joint initiative of the BIS, ECB, Eurostat, IMF, OECD, and the UNSD termed the Statistical Data and Metadata Exchange (SDMX) task force.[6]

---

[5] The ROSC Program is designed to monitor progress in the implementation, and observance, of standards and codes within the IMF's direct operational focus. Reports summarizing countries' observance of these standards are prepared and published on a voluntary basis. They are used to help sharpen policy discussions with national authorities; by rating agencies in their assessments; and in the private sector for risk assessment.

[6] For more information see http://www.sdmx.org.

III. LEVERAGING WEB-BASED TECHNOLOGIES FOR DATA QUALITY ASSESSMENTS

10.   At present, the DSBB does not permit intelligent search and query functions. However, the introduction of a relational database management system (RDBMS) for metadata and the development and rendering of a DSBB model in XML—all part of the ongoing DSBB enhancement project outlined in *Enhancing User Access to Statistical Metadata on the Internet*—would allow for such operations. Under such a regime, DSBB metadata could be interrogated and repurposed to meet specific user needs, including providing information for the DQAF. However, the fact that SDDS metadata provide only about half of the information requirements for a data quality assessment using the DQAF, coupled with the proliferation of statistical metadata on other websites such as national statistical agencies and international/regional organizations, points to the potential for extending the range of such queries beyond the DSBB.

11.   The development and widespread adoption of an XML based open exchange system for the dissemination of statistical information on the Internet, containing standardized vocabularies and schemes to describe metadata, underscores the possibility of data quality assessments based on information derived from an Internet-wide search. Such a system would allow users of the DSBB not only to access but to interrogate and analyze information on other websites using the same standardized vocabularies and schemas for metadata. Therefore, the potential exists to automate the aggregation of information relevant to quality assessments using the DQAF as a set of pre-defined queries. As more organizations adopt the open exchange system for the dissemination of statistical information, both the quantity and quality of such searches should improve.

Table 1. Comparison of Quality Indicators in the DQAF Framework and Metadata in the SDDS Framework.

Data Quality Assessment Framework-Generic Framework			SDDS Framework
Quality Dimensions	**Elements**	**Indicators**	**Metadata**
Prerequisites of quality[7]	**0.1 Legal and institutional environment** – *The environment is supportive of statistics.*	0.1.1 The responsibility for collecting, processing, and disseminating statistics is clearly specified.	"Dissemination of terms and conditions under which official statistics are produced, including ... confidentiality of individual responses" cell.
		0.1.2 Data sharing and coordination among data producing agencies are adequate.	
		0.1.3 Respondents' data are to be kept confidential and used for statistical purposes only.	"Dissemination of terms and conditions under which official statistics are produced, including ... confidentiality of individual responses" cell.
		0.1.4 Statistical reporting is ensured through legal mandate and/or measures to encourage response.	"Dissemination of terms and conditions under which official statistics are produced, including ... confidentiality of individual responses" cell.
	0.2 Resources – *Resources are commensurate with needs of statistical programs.*	0.2.1 Staff, financial, and computing resources are commensurate with statistical programs. 0.2.2 Measures to ensure efficient use of resources are implemented.	
	0.3 Quality awareness – *Quality is a cornerstone of statistical work.*	0.3.1 Processes are in place to focus on quality. 0.3.2 Processes are in place to monitor the quality of the collection, processing, and dissemination of statistics. 0.3.3 Processes are in place to deal with quality considerations, including tradeoffs within quality, and to guide planning for existing and emerging needs.	

---

[7] The elements and indicators included here bring together the "pointers to quality" that are applicable across the five identified dimensions of data quality.

Table 1. Comparison of Quality Indicators in the DQAF Framework and Metadata in the SDDS Framework.

Data Quality Assessment Framework-Generic Framework			SDDS Framework
Quality Dimensions	**Elements**	**Indicators**	**Metadata**
1. Integrity *The principle of objectivity in the collection, processing, and dissemination of statistics is firmly adhered to.*	**1.1 Professionalism** *– Statistical policies and practices are guided by professional principles.*	1.1.1 Statistics are compiled on an impartial basis. 1.1.2 Choices of sources and statistical techniques are informed solely by statistical considerations. 1.1.3 The appropriate statistical entity is entitled to comment on erroneous interpretation and misuse of statistics.	Dissemination of terms and conditions under which official statistics are produced, including ... confidentiality of individual responses" cell.
	1.2 Transparency *– Statistical policies and practices are transparent.*	1.2.1 The terms and conditions under which statistics are collected, processed, and disseminated are available to the public. 1.2.2 Internal governmental access to statistics prior to their release is publicly identified. 1.2.3 Products of statistical agencies/units are clearly identified as such. 1.2.4 Advance notice is given of major changes in methodology, source data, and statistical techniques.	Dissemination of terms and conditions under which official statistics are produced, including ... confidentiality of individual responses" cell. "Identification of internal government access to data before release" "Provision of information about revision and advance notice of major changes in methodology" cell
	1.3 Ethical standards *– Policies and practices are guided by ethical standards.*	1.3.1 Guidelines for staff behavior are in place and are well known to the staff.	

Table 1. Comparison of Quality Indicators in the DQAF Framework and Metadata in the SDDS Framework.

Data Quality Assessment Framework-Generic Framework (July 2001 Vintage)			SDDS Framework
Quality Dimensions	**Elements**	**Indicators**	**Metadata**
2. Methodological soundness *The methodological basis for the statistics follows internationally accepted standards, guidelines, or good practices.*	**2.1 Concepts and definitions –** *Concepts and definitions used are in accord with internationally accepted statistical frameworks.*	2.1.1 The overall structure in terms of concepts and definitions follows internationally accepted standards, guidelines, or good practices: see dataset-specific framework.	"Coverage characteristics" cell, Summary Methodology Statement.
	2.2 Scope – *The scope is in accord with internationally accepted standards, guidelines, or good practices.*	2.2.1 The scope is broadly consistent with internationally accepted standards, guidelines, or good practices: see dataset-specific framework.	"Coverage characteristics" cell, Summary Methodology Statement.
	2.3 Classification/sectorization – *Classification and sectorization systems are in accord with internationally accepted standards, guidelines, or good practices.*	2.3.1 Classification/sectorization systems used are broadly consistent with internationally accepted standards, guidelines, or good practices: see dataset-specific framework.	"Coverage characteristics" cell, Summary Methodology Statement.
	2.4 Basis for recording – *Flows and stocks are valued and recorded according to internationally accepted standards, guidelines, or good practices.*	2.4.1 Market prices are used to value flows and stocks. 2.4.2 Recording is done on an accrual basis. 2.4.3 Grossing/netting procedures are broadly consistent with internationally accepted standards, guidelines, or good practices.	"Coverage characteristics" cell, Summary Methodology Statement. "Coverage characteristics" cell, Summary Methodology Statement. "Coverage characteristics" cell, Summary Methodology Statement.

Table 1. Comparison of Quality Indicators in the DQAF Framework and Metadata in the SDDS Framework.

Data Quality Assessment Framework-Generic Framework (July 2001 Vintage)			SDDS Framework
Quality Dimensions	**Elements**	**Indicators**	**Metadata**
3. Accuracy and reliability **Source data and statistical techniques are sound and statistical outputs sufficiently portray reality.**	**3.1 Source data** – *Source data available provide an adequate basis to compile statistics.*	3.1.1 Source data are collected from comprehensive data collection programs that take into account country-specific conditions. 3.1.2 Source data reasonably approximate the definitions, scope, classifications, valuation, and time of recording required. 3.1.3 Source data are timely.	"Coverage characteristics" cell, Summary Methodology Statement. Summary Methodology Statement.
	3.2 Statistical techniques – S*tatistical techniques employed conform to sound statistical procedures.*	3.2.1 Data compilation employs sound statistical techniques. 3.2.2 Other statistical procedures (e.g., data adjustments and transformations, and statistical analysis) employ sound statistical techniques.	Summary Methodology Statement. Summary Methodology Statement.
	3.3 Assessment and validation of source data–*Source data are regularly assessed and validated.*	3.3.1 Source data—including censuses, sample surveys and administrative records—are routinely assessed, e.g., for coverage, sample error, response error, and non-sampling error; the results of the assessments are monitored and made available to guide planning.	
	3.4 Assessment and validation of intermediate data and statistical outputs.- *Intermediate results and statistical outputs are regularly assessed and validated.*	3.4.1 Main intermediate data are validated against other information where applicable. 3.4.2 Statistical discrepancies in intermediate data are assessed and investigated. 3.4.3 Statistical discrepancies and other potential indicators of problems in statistical outputs are investigated.	"Dissemination of ... that support statistical cross-checks and provide assurance of reasonableness" cell. "Dissemination of ... that support statistical cross-checks and provide assurance of reasonableness" cell. "Dissemination of ... that support statistical cross-checks and provide assurance of reasonableness" cell.
	3.5 Revision studies – *Revisions, as a gauge of reliability, are tracked and mined for the information they may provide.*	3. 5.1 Studies and analyses of revisions are carried out routinely and used to inform statistical processes.	

Table 1. Comparison of Quality Indicators in the DQAF Framework and Metadata in the SDDS Framework.

Data Quality Assessment Framework-Generic Framework (July 2001 Vintage)			SDDS Framework
Quality Dimensions	**Elements**	**Indicators**	**Metadata**
4. Serviceability **Statistics are relevant, timely, consistent, and follow a predictable revisions policy.**	**4.1 Relevance –** *Statistics cover relevant information on the subject field.*	4.1.1 The relevance and practical utility of existing statistics in meeting users' needs are monitored.	
	4.2 Timeliness and periodicity – *Timeliness and periodicity follow internationally accepted dissemination standards.*	4.2.1 Timeliness follows dissemination standards. 4.2.2 Periodicity follows dissemination standards	"Timeliness" cell. "Periodicity" cell.
	4.3 Consistency – *Statistics are consistent within the dataset, over time, and with major datasets.*	4.3.1 Statistics are consistent within the dataset (e.g., accounting identities observed). 4.3.2 Statistics are consistent or reconcilable over a reasonable period of time. 4.3.3 Statistics are consistent or reconcilable with those obtained through other data sources and/or statistical frameworks.	"Dissemination of ... that support statistical cross-checks and provide assurance of reasonableness" cell.
	4.4 Revision policy and practice – *Data revisions follow a regular and publicized procedure.*	4.4.1 Revisions follow a regular, well-established and transparent schedule. 4.4.2 Preliminary data are clearly identified. 4.4.3 Studies and analyses of revisions are made public.	"Provision of information about revision and advance notice of major changes in methodology" cell. "Provision of information about revision and advance notice of major changes in methodology" cell. "Provision of information about revision and advance notice of major changes in methodology" cell.

Table 1. Comparison of Quality Indicators in the DQAF Framework and Metadata in the SDDS Framework.

Data Quality Assessment Framework-Generic Framework (July 2001 Vintage)			SDDS Framework
Quality Dimensions	**Elements**	**Indicators**	**Metadata**
5. Accessibility *Data and metadata are easily available and assistance to users is adequate.*	**5.1 Data accessibility –** *Statistics are presented in a clear and understandable manner, forms of dissemination are adequate, and statistics are made available on an impartial basis.*	5.1.1 Statistics are presented in a way that facilitates proper interpretation and meaningful comparisons (layout and clarity of text, tables, and charts). 5.1.2 Dissemination media and formats are adequate. 5.1.3 Statistics are released on the pre-announced schedule. 5.1.4 Statistics are made available to all users at the same time. 5.1.5 Non-published (but non-confidential) sub-aggregates are made available upon request.	"Dissemination formats" page. "Advance Release Calendar" cell, "Advance Release Calendar" page and National Summary Data Pages (NSDP). "Simultaneous release to all interested parties" cell.
	5.2 Metadata accessibility – *Up-to-date and pertinent metadata are made available.*	5.2.1 Documentation on concepts, scope, classifications, basis of recording, data sources, and statistical techniques is available, and differences from internationally accepted standards, guidelines or good practices are annotated. 5.2.2 Levels of detail are adapted to the needs of the intended audience.	"Dissemination of documentation on methodology and sources used in preparing statistics" cell. "Coverage characteristics" and "Dissemination of ... that support statistical cross-checks and provide assurance of reasonableness" cell.
	5.3 Assistance to users – *Prompt and knowledgeable support service is available.*	5.3.1 Contact person for each subject field is publicized. 5.3.2 Catalogues of publications, documents, and other services, including information on any charges, are widely available.	"Contact person" page.

DIFFERENCES IN PURPOSE AND SCOPE IN EACH OF THE SDDS/DQAF DIMENSIONS

12.   ***Data*** *(in the SDDS)/**Serviceability***(*in the DQAF*)**.** The SDDS prescribes the data to be disseminated (18 data categories and certain breakdowns) along with the periodicity and timeliness for each category. The DQAF does not prescribe coverage; rather the DQAF identifies, among the pointers to be considered in assessing quality, the existence of a process to monitor the relevance and practical utility of the existing data in meeting users' needs. For periodicity and timeliness, the DQAF identifies, as a pointer to good practice, observance of the Fund's dissemination standards. In addition, the DQAF identifies, within its serviceability dimension, practices with respect to consistency and revisions that are in other SDDS dimensions.

13.   ***Access by the public/Accessibility.*** The SDDS prescribes two practices; the DQAF identifies the same two practices as pointers to good practice. In addition, the DQAF identifies as relevant to assessing quality additional practices about data accessibility, practices about metadata accessibility (see below about the SDDS quality dimension), and practices about assistance to users.

14.   ***Integrity/Integrity***. The SDDS prescribes dissemination of information about four key practices. The DQAF identifies as relevant to assessing quality the same indicators of integrity, all grouped together under the heading of transparency. In addition, the DQAF identifies as relevant in assessing data quality indicators related to professionalism and ethical standards. The SDDS prescription about dissemination of information about revisions has a counterpart in the DQAF within serviceability.

15.   ***Quality/Methodological soundness, Accuracy and reliability***. The SDDS prescribes dissemination of information to help users make their own assessment of data quality. The summary methodologies that are called for by the SDDS provide a structure designed to elicit key information needed to do this. The DQAF identifies the practices of disseminating this information as relevant in assessing quality (within its accessibility and serviceability dimensions). In addition and importantly, the DQAF details practices about data collection and production that shed light on quality by providing two separate dimensions—methodological soundness, and accuracy and reliability. The resemblance between the six items of the SDDS summary methodologies and the elements of these two DQAF dimensions of quality is worth noting. For example, the first bullet item listed in the SDDS summary methodologies—analytical framework, concepts, and definitions—has a close counterpart as item 2.1—concepts and definitions—in the DQAF dimension of methodological soundness.