

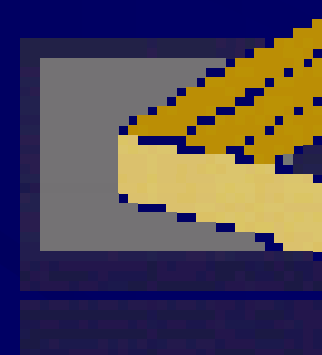
Joint UNECE/EUROSTAT Work Session on Statistical Metadata

The Use of Metadata in Data Quality Assessments: An SDDS Example

**Robert Di Calogero, Paul Austin, Gary Jones, and Michel Le Marois
Statistics Department, International Monetary Fund**



Luxembourg
March 6–8, 2002



Objective

1. 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.
2. 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.

The SDDS Metadata Framework and Data Quality Assessments

- Since the launching of the DSBB in September 1996, subscription to the SDDS has grown to 50 countries .
- The European Central Bank and Eurostat have adopted the SDDS format for their metadata presentations, and have provided links to the DSBB.
- Metadata for 39 participants in the IMF's General Data Dissemination System (GDDS) built on the same four dimensions—data characteristics, quality, access, and integrity—are posted on the DSBB.

The SDDS Metadata Framework and Data Quality Assessments

- 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.

Data Quality Assessment Framework

- 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).
- The DQAF is intended as a tool to assess the quality of the collection, production, and dissemination of data.

Coverage of SDDS Metadata vis-à-vis the DQAF Dimensions

- 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.

Coverage of SDDS Metadata vis-à-vis the DQAF Dimensions

- 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.

Coverage of SDDS Metadata vis-à-vis the DQAF Dimensions

- 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.

Coverage of SDDS Metadata vis-à-vis the DQAF Dimensions

- 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.

Leveraging Web-based Technologies for Data Quality Assessments

- At present, the DSBB does not permit intelligent search and query functions.
- 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.

Leveraging Web-based Technologies for Data Quality Assessments

- The development and widespread adoption of an XML based open exchange system for the dissemination of statistical information on the Internet, containing standardized vocabularies to describe metadata, underscores the possibility of data quality assessments based on information derived from an Internet-wide search.
- 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.