Using Metadata Standards to Improve National and IMF DATA

Nalini Umashankar
International Monetary Fund (IMF), USA
numashankar@imf.org

Keywords: data quality; International Monetary Fund (IMF).

1. Introduction[1]

Metadata standardization leads to greater efficiencies and lower costs in global exchange and internal production of data. Use of metadata standards enhances the accountability of countries for providing quality information about their economy and improves the understanding of data by users. The IMF experience, as outlined in this abstract, demonstrates how metadata standards have resulted in faster, cheaper and more consistent production and dissemination of data.

2. Using Metadata to Monitor Data Quality

One of the challenges that we face is to best guide member countries to provide quality information about their economy, while holding them accountable for dissemination of such information. In this regard, the IMF has established the Data Standards Initiatives[2] for guiding the dissemination statistical data and metadata. They comprise the General Data Dissemination System, the Special Data Dissemination Standard (SDDS) and the SDDS Plus. Countries following these standards use the IMF Data Quality Assessment Framework (DQAF). The framework defines five data quality aspects:

1. Integrity
2. Methodological Soundness
3. Accuracy and Reliability
4. Serviceability
5. Accessibility

Countries are required to submit text (metadata) describing their data and dissemination practices to the IMF for posting on an electronic bulletin board. Each country is responsible for the accuracy and timeliness of its metadata. IMF Staff review countries' postings to ensure they provide comprehensive and internationally comparable metadata.

The benefits of the Dissemination Standards are that countries are made accountable for providing metadata about their data and users are able to obtain consistent information that facilitates cross- country comparisons as well as a better understanding of the quality of a country's statistics.

The generic DQAF framework serves as an umbrella for these dataset-specific frameworks: National accounts, Consumer prices, Producer prices, Government Finance Statistics and Public Sector Debt Statistics, Monetary Statistics, Balance of Payments and International Investment Position Statistics and External Debt.[3]

[1] The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.
[2] For more information on Dissemination Standards at the IMF, refer to the following link: http://dsbb.imf.org/
[3] For further information, refer to the link: http://dsbb.imf.org/Pages/DQRS/DQAF.aspx

Content of the Framework

The elements and indicators within their respective dimensions are as follows:

0 Prerequisites of quality: Although it is not a dimension of quality, this group of "pointers to quality" includes elements and indicators that have an overarching role as prerequisites, or institutional preconditions, for quality of statistics. Note that the focus is on the agency, such as a national statistical office, central bank, or a ministry/department. These prerequisites cover the following elements:

0.1 Legal and institutional environments
0.2 Resources available for the statistical program
0.3 Relevance
0.4 Other quality management

1 Assurances of integrity: This dimension refers to the adherence to the principle of objectivity in the collection, compilation, and dissemination of statistics. The dimension encompasses institutional arrangements that ensure professionalism in statistical policies and practices, transparency, and satisfactory ethical standards. The three elements for this dimension of quality are the following:

1.1 Institutional Integrity
1.2 Transparency
1.3 Ethical standards

2 Methodological soundness: This dimension specifies that the methodological basis for the production of statistics should be sound and that this can be attained by following internationally accepted standards, guidelines, or good practices. This dimension is necessarily dataset-specific, reflecting different methodologies for different datasets. This dimension has four elements, namely:

2.1 Concepts and definitions
2.2 Scope
2.3 Classification/sectorization
2.4 Basis for recording

3 Accuracy and reliability: This dimension says that statistical outputs should sufficiently portray the reality of the economy. This dimension is also data specific, reflecting the sources used and their processing. The five elements of this dimension cover the following:

3.1 Source data
3.2 Assessment of source data
3.3 Statistical techniques
3.4 Assessment and validation of intermediate data and statistical outputs.
3.5 Revision studies

4 Serviceability: This dimension indicates that statistics are disseminated with an appropriate periodicity in a timely fashion, are consistent internally and with other major datasets, and follow a regular revision policy. The three elements for this dimension are as follows:

4.1 Periodicity and timeliness
4.2 Consistency
4.3 Revision policy and practice

5 **Accessibility**: This dimension specifies data and metadata to be presented in a clear and understandable manner on an easily available and impartial basis, that metadata are up-to-date and pertinent, and that a prompt and knowledgeable support service is available. This dimension has three elements:

5.1 Data accessibility
5.2 Metadata accessibility
5.3 Assistance to users

3. Using Metadata to Exchange Data

A second challenge is how to make the global exchange of data and metadata faster and cheaper, while reducing the reporting burden of member countries. The *Statistical Data and Metadata Exchange (SDMX)*[4] standards enable us to meet this challenge. IMF uses SDMX to reduce the work to map metadata structures between agencies, by collaborating internationally to develop common *Data Structure Definitions (DSDs)*, code lists and concept schemes. The IMF increasingly collect and disseminate data using SDMX.

SDMX promises to make global data exchange faster, cheaper, and more consistent, thereby reducing the reporting burden on member countries. For example, the external sector DSD developed by five SDMX sponsoring organizations, under the leadership of the IMF, will allow country authorities to provide one SDMX data file or web service to satisfy the data collection needs of multiple international institutions. Countries will report external sector data using the DSD and standard codes to an agency or a registry, and consuming organizations will process data from files that could be shared across processing organizations or pulled from a data providers' web service using information provided to a registry.[5]

4. Using Metadata to Improve Data Consistency

A third challenge is how to improve data consistency by using consistent names and definitions in data production and dissemination systems. The IMF uses master lists for economic concepts and country names to improve data comparisons and make data processes more efficient. The *Catalog of Time Series* (CTS) provides a hierarchical vocabulary of concepts across economic sectors, and contains names and codes for over 50,000 items. We also maintain standard country names and codes. The Statistics Department manages the master lists in collaboration with other Departments. IMF databases use these master lists in their production systems. Many IMF applications of SDMX use the Catalog of Time Series.

Using master lists ensures the consistent use of names and definitions across databases. Managing these lists centrally allows accuracy and methodology checks to be done once rather than multiple times. Using these lists, with common definitions, in all systems makes it very easy to compare data across countries and economic concepts. The use of these master lists in SDMX leads to faster and cheaper exchange of data and metadata.

D. Using Metadata to Explain Data Characteristics

A fourth challenge is how to improve users' understanding of data and methodology used in compilation of statistics. To this end, IMF published outputs contain explanatory notes, which help users understand the data they see. We store these explanatory notes (which we call "referential metadata)" in a standard format using a DQAF-based schema, allowing us to present these notes in the way most suited to any given output format.

[4] For more information, see: link http://www.sdmx.org/
[5] A registry is a web-based application that enables reporting countries to access the DSDs, code lists and report data in SDMX format.

The benefit of using standards in presentation of metadata is that users are able to view metadata alongside relevant data and understand the characteristics of data better; it facilitates cross country comparisons of data and enables structuring content using a comprehensive framework.

Conclusion

To summarize, the poster illustrates how metadata standardization in the IMF leads to an improvement in the quality of statistical information and a better understanding of data and metadata by users. It enables a more efficient and faster exchange of information at lower costs, which is made possible as a result of collaboration with member countries and other international organizations.