

UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION  
AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE

Work Session on Statistical Metadata (METIS)  
(Geneva, Switzerland, 10-12 March 2010)

## RECENT DEVELOPMENTS IN METADATA AT THE INTERNATIONAL MONETARY FUND<sup>1</sup>

Nalini Umashankar, IMF

### I. INTRODUCTION

1. This paper describes the significant investments in metadata made by IMF for developing standards with the implementation of SDMX, the Data Warehouse (IMF.Stat), Metastore, and Schemalogic. It also describes the initiatives for repurposing metadata of the IMF Dissemination Standards Bulletin Board (DSBB) in publications and enhancements in metadata collection in the DSBB, along with issues concerning metadata management within the Fund.

2. *There is a growing emphasis on metadata driven by an increasing need to understand data and the need for standardization in data sharing.* In the area of collection, the establishment of the DSBB required countries to report metadata to comply with the requirements of the IMF Data Standards' Initiatives. The introduction of SDMX for collection of some data categories strengthened the need for standardization of metadata. The recent introduction of the IMF.Stat and Metastore has highlighted the importance of metadata for an improved understanding of data from different sources. The IMF country desk database, also called Operational Database of the Fund, has required developing a mechanism for storing metadata which flows into the IMF.Stat to achieve harmonization. The establishment of a central repository for metadata, Schemalogic, has led to the need for greater standards in developing and

---

<sup>1</sup> The views expressed in this paper are those of the author and should not be attributed to the IMF, its Executive Board, or its Management.

maintaining metadata. All of these initiatives have surfaced issues relating to metadata management which are being addressed.

## Structural Metadata

### SDMX developments

3. *Developments in SDMX have led to standardization in structural Metadata.*<sup>2</sup> The Data Structure Definition (DSD) for Balance of Payments (BOP) Statistics implemented by the IMF in collaboration with international organizations is an important initiative for adopting a standard structure for structural metadata in an SDMX data exchange. BOP data are collected via the Integrated Correspondence System (ICS), a secure web application, with a DSD that conforms to international SDMX standards. There is exchange of SDMX data with international organizations like OECD and FAO making use of the SDMX web services implemented to the IMF.Stat. There is a new project to develop a DSD for government Finance Statistics. In the area of data collection, the IMF has been trying to encourage countries to report using SDMX, though the momentum has been slow. The technical infrastructure has been developed for data collection and data exchange. One of the upcoming projects is the development of a Data Structure Definition (DSD) for Principal Global Indicators, an inter-agency website, which will support download and collection of data in SDMX format.

### IMF.Stat

4. *With the establishment of the enterprise economic data warehouse, IMF.Stat, the importance of structural and referential metadata has been strengthened.* The data warehouse, IMF.Stat, has been developed in collaboration with the OECD. The IMF.Stat uses six common dimensions to store data.

- *Country and Group* – an authoritative set of country names and geographic and analytical groupings of countries.
- *Economic Concept* – listing of economic concepts defined in the Fund’s Catalog of Economic Time Series as the authoritative source.
- *Data Source* – authoritative listing of the source used for the data.
- *Unit* – currently treated as an aggregate of unit of measure with scale, e.g., millions of U.S. dollars, to be separated in the near future.
- *Status* – an authoritative listing of values to manage access to data .e.g., shareable.
- *Time, Frequency* – authoritative lists of values related to observation intervals.

5. Where possible, the intention is to store all data in a single dataset, using these common dimensions to facilitate comparison of data across Data Source and other dimensions. Use of a single dataset implies the need for standardization of structural metadata.

---

<sup>2</sup> In SDMX, a distinction is made between “structural” metadata that define the structure of statistical data sets and “referential” metadata describing actual metadata contents, for instance, concepts and methodologies used, the data quality (e.g., accuracy and timeliness) and the production and dissemination process (e.g., contact points, release policy, dissemination formats).

6. A feature of IMF.Stat is that it provides the ability to store referential metadata with the data. This is seen as a major benefit as it allows the user to easily access information relevant to the data.

**Figure 1. Referential Metadata Available From IMF.Stat Red “i”**

The screenshot displays the IMF.Stat web interface. At the top, there is a navigation bar with the IMF STAT logo, a welcome message for 'Nalini Umashankar', and links for 'Contact Us', 'User Guide', and 'Home'. Below this, a breadcrumb trail shows the current data selection: 'Country & Group [1 / 491] Economic Concept [2075 / 5175] Data Source [22 / 29] Status [1 / 1]'. A toolbar with various icons is visible below the breadcrumb. The main content area shows a table of data extracted on 2010/01/21 11:33 from IMF.Stat, with page 1-42 of 513. The table has columns for Country & Group, Data Source, Status, Frequency, and Time. A metadata popup window is open over the table, titled 'Metadata' and showing details for 'Country and Group: United States Concept: Industrial Production'. The popup includes sections for Source (Direct source: Board of Governors of the Federal Reserve System (USA)), Concepts & Classifications (Weights: Weights reference period: 1997; Basis for Calculation: the weights are based on annual estimates of value added), and Population & Scope (Sector coverage: The index covers manufacturing, mining, and electric and gas utilities; Generally, the coverage of industrial production indices comprises mining and quarrying, manufacturing and electricity, and gas and water, according to the UN International Standard Industrial Classification (ISIC). IFS includes data on National Accounts, Indicators of Economic Activity, Labor Markets, Prices, Government and Public Sector Finance, Financial Indicators, Balance of Payments, International Investment Position, International Reserves, Fund Accounts, External Trade).

## II. MAINTENANCE OF STRUCTURAL METADATA

### Metastore

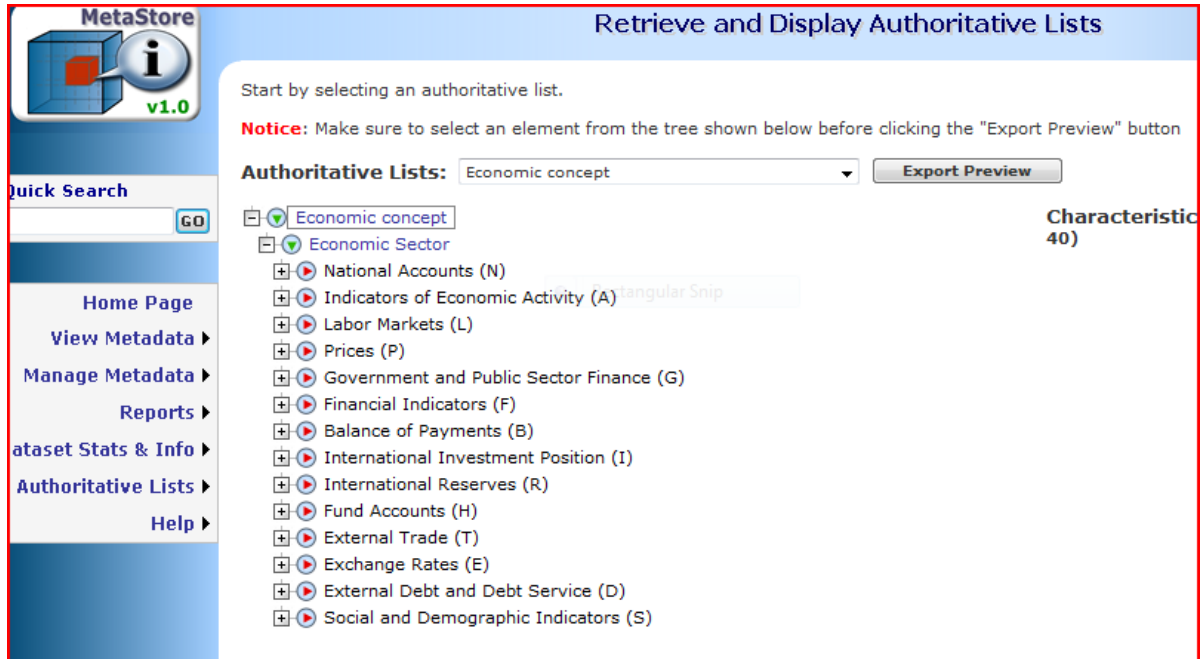
7. *Metastore provides a platform for managing structural and referential metadata for the IMF.Stat.* The authoritative lists for the six common dimensions in the IMF.Stat are maintained in the Metastore, which is a central repository for metadata, developed in collaboration with the OECD.

8. MetaStore has the following features

- the central repository of both structural and referential metadata;
- the source of all metadata used in IMF.Stat;
- the store of mappings that relate production metadata to the authoritative source;
- an inclusion of the forty-two (42) metadata types used by the OECD;
- an ability to attach metadata to multiple levels of data (data set, country, concept, individual observation.)

9. Efforts are in progress to align the authoritative lists to the SDMX Cross-Domain concepts and code lists, wherever possible.

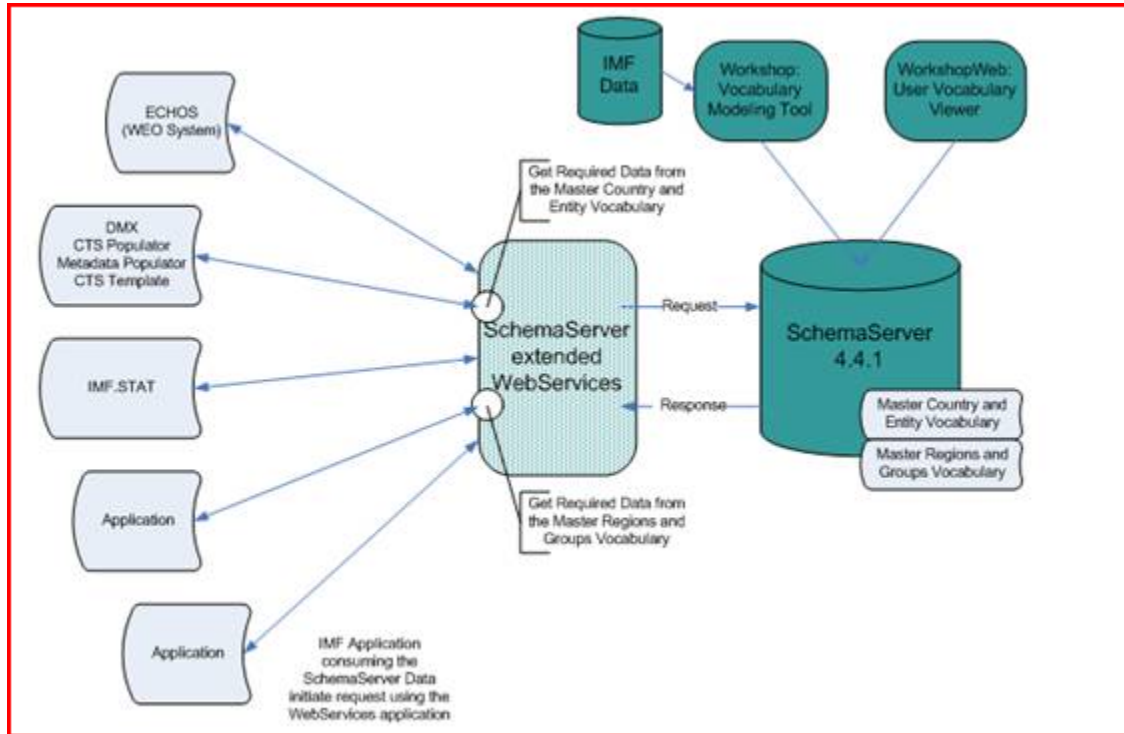
**Figure 2. Metastore: The Current Tool to Manage Structural and Referential Metadata**



## Schemalogic

10. ***Schemalogic will be used to manage business vocabularies in the IMF.*** The Fund is in the process of moving the authoritative lists from Metastore to SchemaLogic Enterprise Suite, an enterprise tool for managing metadata and business vocabulary. With the move to Schemalogic, there will be a single flow of structural metadata to the IMF.Stat and production systems in the Fund. The economic concept vocabulary and the country/region vocabulary have been populated in Schemalogic. The technical functionalities are being developed for use of the tool on a daily basis. When the business needs are addressed, all authoritative lists will be stored in one place and flow to production systems and the IMF.Stat via web services. However, referential metadata will continue to reside in Metastore.

**Figure 3. Schemalogic: the Future Tool to Manage Authoritative Lists**



## Referential Metadata

### DSBB enhancements

11. ***The DSBB is an important initiative providing access to referential metadata from member countries and metadata management is being automated.*** The DSBB provides access to the referential metadata that are prepared by the subscribers to the IMF Special Data Dissemination Standard (SDDS) and participants to General Data Dissemination System (GDDS) for the data categories these data standards' initiatives comprise. The use of Data Quality Assessment Framework (DQAF) for SDDS and GDDS countries helps in standardizing referential metadata. Currently there are 67 SDDS subscribing countries and 98 GDDS participating countries. There are recent enhancements in metadata processing for the DSBB. The collection of metadata will be handled using the Integrated Correspondence System (ICS), a secure web application. The countries will report metadata to the IMF using the ICS web application. With the move to the new system, referential metadata will be updated in an automated way. The countries will validate the accuracy of metadata once a year using the online Certification Editor. The IMF is looking into an expansion in the number of data categories and a better presentation of the international methodologies in the country metadata in the DSBB.

12. There are efforts undertaken to streamline the collection of Balance of Payments metadata for the DSBB and for dissemination by the Statistics Department. This may lead to largely repurposing for publications the DSBB referential metadata supplemented by more information. From the perspective of data dissemination in the IMF *International Financial Statistics (IFS)* and IMF.Stat, this repurposing would eliminate the current metadata collection and manual updating processes that exist for these

dissemination media. These processes would be replaced by an automated loading of metadata by mapping the important DQAF metadata categories to the metadata types in Metastore. It would also lead to reduction in the reporting burden for countries, if implemented.

### III. REFERENTIAL METADATA IN METASTORE

13. *Metastore facilitates classification of referential metadata and feeds into IMF.Stat.* The IMF and OECD use Metastore for storing referential metadata. While the OECD uses 42 metadata types for classifying metadata, the IMF uses a subset of these types. In addition some of the types such as agency information are a predefined field and not free form in the IMF implementation. There is more granularity in the attachment of metadata as they can be attached at the country/concept, concept, dataset and observation levels.

14. Referential metadata are presented at various attachment levels in Metastore and IMF.Stat. In addition to referential metadata, breaks in series are represented by symbols next to data in IMF.Stat with legends explaining the breaks. Efforts are being made to align the metadata types to the SDMX cross-domain concepts, wherever possible.

15. The latest version of the Principal Global Indicators (PGI) inter-agency website (available at [www.principalglobalindicators.org](http://www.principalglobalindicators.org)) provides referential metadata at granular levels. This is important to explain differences in data between different organizations participating in the initiative.

**Figure 4. Metastore and Attachment Levels for IMF.Stat Metadata**

The screenshot displays the 'Coordinate and New Metadata Management' page in the Metastore application. The page title is 'Coordinate and New Metadata Management' and it includes a navigation menu on the left with options like 'Quick Search', 'Home Page', 'View Metadata', 'Manage Metadata', 'Reports', 'Dataset Stats & Info', 'Authoritative Lists', and 'Help'. The main content area shows a list of coordinates with a search box containing '914.....'. Below the search box, there are several management utilities: 'Edit Selected Coordinate Descriptions', 'View Selected Coordinate Metadata', 'Batch Process Selected Coordinate', and 'Remove Selected Coordinate from List'. A section titled 'Add new Metadata to the Coordinates listed above:' (with a note that items marked with \* are new) contains a list of 'Metadata elements\*'. These elements include 'IMF.Stat Mapping elements' and 'Types of Metadata', which is further broken down into 'Source', 'Data characteristics', 'Statistical population and scope of the data', 'Statistical concepts and classifications used', 'Manipulation and dissemination', and 'Other Aspects'.

### IV. ISSUES BEING CONSIDERED FOR METADATA MANAGEMENT

16. *Metadata management is gaining importance with developments in the data management initiatives in the Fund.* The issue of governance of structural metadata is being discussed along with

mechanisms to strengthen the content of the authoritative lists, in particular, of the economic concept vocabulary. With the establishment of an owner of the lists, there will be progress toward higher standards in maintenance of structural metadata including closer alignment with the SDMX standards. The IMF country desks are aligning their databases with the structural metadata vocabularies with a move toward structured databases.

17. As regards referential metadata, the medium to long term strategy is to use the DSBB DQAF metadata and supplement it with additional information. In the interim, existing metadata is being attached at granular levels in Metastore and IMF.Stat. Outside of the Statistics Department, owners of databases in IMF are recognizing the importance of metadata by creating additional metadata fields in their databases to store referential metadata which then flow into the IMF.Stat, which is the enterprise economic data warehouse.

## GLOSSARY

**IMF.Stat:** An enterprise economic data warehouse of the IMF developed in collaboration with OECD.

**Metastore:** A central repository of structural and referential metadata developed in collaboration with OECD.

**Schemalogic:** An enterprise tool for storing business vocabularies.

**DSBB:** Dissemination Standards Bulletin Board <http://www.dsbb.imf.org>

**SDDS:** Special Data Dissemination Standards.

**GDDS:** General Data Dissemination System.

**DQAF:** Data Quality Assessment Framework. Five dimensions—assurances of integrity, methodological soundness, accuracy and reliability, serviceability, and accessibility—of data quality and a set of prerequisites for data quality are the center of the IMF Data Quality Assessment Framework (DQAF). The DQAF, which is used for comprehensive assessments of countries' data quality, covers institutional environments, statistical processes, and characteristics of the statistical products.

**SDMX:** Statistical Data and Metadata Exchange

**DSD:** Data Structure Definition, used for exchange of datasets in SDMX

**Operational Database of the Fund (ODF):** Dataset containing IMF country desk data.