STATISTICAL COMMISSION and   **COMMISSION OF THE**
ECONOMIC COMMISSION FOR EUROPE   **EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS   **EUROSTAT**

Joint UNECE/Eurostat Work Session   Working Paper No. 3
on Statistical Metadata   English only
(6 - 8 March 2002, Luxembourg)

Topic (i): Infrastructure issues for statistical metadata

ENHANCING USER ACCESS TO STATISTICAL METADATA ON THE INTERNET

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

Invited paper

I.   INTRODUCTION

1.   As part of an ongoing effort to enhance the functionality and user-friendliness of the Dissemination Standards Bulletin Board (DSBB), the IMF is implementing an open exchange system for the dissemination and exchange of statistical information on the internet. This system will leverage the success of the DSBB in disseminating metadata in a widely recognized format – sometimes referred to as the SDDS format. The system also will increase the value of the investments made by SDDS subscribers and GDDS participants in developing the metadata posted on the DSBB, as well as the wealth of statistical data and metadata that is linked to the website.

2.   Key elements of the open exchange system being developed by the IMF are the DSBB metadata model – and its rendering in XML - and the corresponding Relational Database Management System (RDBMS) for metadata. The RDBMS will reflect the structure of both the SDDS and GDDS metadata. The enhanced system will provide for more sophisticated and reliable data management (storage/retrieval) functions, while supporting improved information processing and analytical capabilities for DSBB content, such as, for example, intelligent searching and dynamic querying, information discovery functions, as well as the use of content aggregation technology to automate SDDS observance monitoring. These enhancements and their application to a search and query facility are discussed below.

II.   THE DSBB METADATA MODEL

3.   The inclusion of a complete set of metadata on Eurostat's Euro Indicators website in SDDS format – harmonized with metadata in SDDS format previously developed by the ECB and posted on the DSBB - has furthered the IMF's objective of providing access to metadata and associated data sets that complement and enhance the information on the DSBB. [2] These links have strengthened the association of DSBB metadata with actual data beyond those data series available through hyperlinks to SDDS subscribers' National Summary Data Pages. Moreover, the similarity in presentation format for statistical metadata on these sites enables data users worldwide to gain access to information from multiple sites in a readily recognizable and comparable form.

---

4.   Nevertheless, despite existing commonalities in the format in which metadata are disseminated, it is not yet possible to perform intelligent searches and queries - either free-form or predefined – of the DSBB or other websites using this format. The IMF is working to address this need by building the portal and data mining capabilities necessary to offer users the ability to search and query content that is both internal and external to the DSBB.

5.   As a first step, Fund staff have developed a DSBB metadata model that encompasses both SDDS and GDDS metadata.The model captures essential information on a) the coverage, periodicity, and timeliness of the data; b) access to the data including, for SDDS subscribers, advance release calendars; c) the integrity of the data including the terms and conditions under which the data are compiled and disseminated; and d) quality such as the identification of methodological publications and related data series that serve as a cross check on the data. In addition, the model encompasses summary methodology statements, prepared by SDDS subscribers according to a standard outline, that provide a more detailed description of data compilation practices.[1] Other elements captured by the model include the comparison of SDDS subscribers' actual data on National Summary Data Pages against the description of that data on the DSBB and GDDS participants' plans for statistical improvement.

6.   Diagrams A & B schematically present the DSBB metadata model from two perspectives. Diagram A presents a "high-level" view representing the first breakdown for each of the four dimensions (data, access, integrity, and quality) plus summary methodology statements, SDDS observance, and GDDS plans for improvement. Diagram B shows the breakdowns of the metadata elements for summary methodology statements. It should be noted that the highest level view (Analytical Framework, Concepts, Definitions, etc.) for summary methodology statements corresponds to the breakdown for coverage under the data dimension of the model, thereby allowing searches over both the base page and summary methodology page metadata.

7.   The content of the metadata model for each country whose metadata are posted on the DSBB will be stored in a relational database (RDBMS). This will permit sophisticated search and query operations on the metadata and provide the backbone for a content management system (CMS) to manage DSBB metadata on the web. The DSBB website will be transformed from a set of static pages to one in which pages are dynamically generated on request, thereby increasing its flexibility to meet specific user needs. DSBB users will be able to order up a set of web pages containing information gleaned from a search and query operation in a format tailored to their purpose. Moreover, rendering of the DSBB metadata model in Extensible Markup Language (XML) will facilitate search and query operations on metadata on other sites using the same open exchange system.

III.   LEVERAGING XML TECHNOLOGIES

8.   Extensible Markup Language (XML) is being used by the IMF to render the metadata model. XML provides the syntax and tagged format to describe metadata, and metadata language vocabularies, which will be organized in terms of a set of structured elements, or Document Type Definitions (DTDs). This will enable DSBB metadata to be exchanged, published, and disseminated in a more effective manner and in a multitude of formats without having to change the underlying metadata descriptions. Additionally, working in conjunction with the CMS, the XML tagged metadata content can be effectively managed (i.e., edited, validated, stored, retrieved, etc.). Moreover, users on other websites, such as country-specific sites or third-party information aggregators, may access the metadata and repurpose it for their own use.

9.   Leveraging XML to render statistical metadata on the web is an important step forward in building an open exchange system for the dissemination of statistical data and metadata over the internet.

---

[1] It should be noted that the information provided by SDDS subscribers in the context of the "Summary Methodology" statements has a counterpart – albeit with a lesser degree of detail – in the "Comprehensive Frameworks" (Table A) of the metadata provided by GDDS participants.

This work is being undertaken in the context of the Statistical Data and Metadata Exchange (SDMX) initiative.[1]

IV. THE DSBB SEARCH AND QUERY FACILITY

10. The IMF is in the process of developing the DSBB as a portal facility linking SDDS/GDDS metadata with metadata and data disseminated on other national, international, and regional organizations' websites. This endeavour, as well as the ability of other websites to link to the DSBB, will be enhanced by the development of an effective XML based search and query facility on the DSBB. The IMF also will be looking at ways to allow users of the DSBB to search and query metadata on other websites which employ the same metadata model rendered in XML.
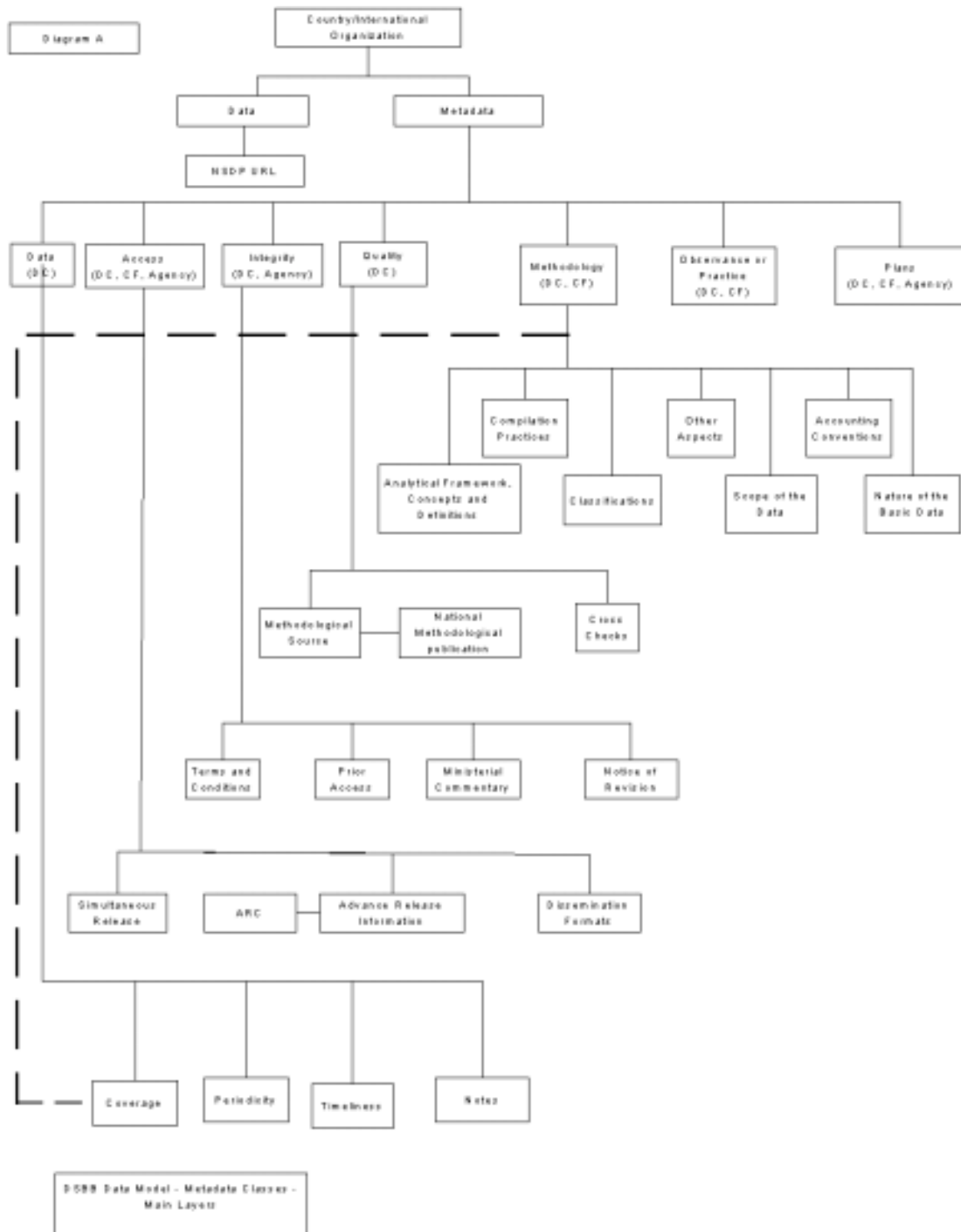
11. An example of how the search and query facility would operate on the DSBB would be as follows: At present, obtaining basic information on the Producer Price Index (PPI) across all SDDS subscribers (e.g., index base year, number of establishments covered, number of commodities covered) requires a review, on a page by page basis, of both the metadata Base Page as well as the summary methodology statement for each subscriber. Moreover, additional interrogation of the metadata in the form of a query across both subscribers and data categories, e.g., identifying whether the sample of establishments used to compile the PPI is identical to that used in compiling the Industrial Production Index, would require reviewing two complete sets of metadata for each subscriber.

12. The problem discussed above—the tedium of searching multiple pages for some 50 subscribers—could be addressed through the use of the search and query facilities. Tables 1 – 4 illustrate the approach being developed for the DSBB. Table 1 shows how users will be provided a set of criteria to establish the parameters of the query to be performed (countries, economic sector, data category, metadata class, and metadata elements). Table 2 illustrates the PPI query discussed above while Tables 3 and 4 represent the results of the query and the possibility to further refine the search criteria, respectively. Under the system being developed query results could be displayed in a multitude of formats, and the results could be passed along to other applications.

13. The enhancements discussed above will enable the DSBB to meet new and changing requirements for disseminating metadata. In the coming months, as these enhancements to the DSBB are implemented, IMF staff look forward to working collaboratively with other organizations interested in adapting the DSBB metadata model rendered in XML for the dissemination of statistical information on the Internet.

---

[1] For additional information see http://www.sdmx.org

Diagram A

Country/International Organization

- **Data**
  - **NSDP URL**
- **Metadata**
  - **Data** (DC)
  - **Access** (DC, CF, Agency)
  - **Integrity** (DC, Agency)
  - **Quality** (DC)
  - **Methodology** (DC, CF)
    - **Compilation Practices**
    - **Analytical Framework, Concepts and Definitions**
    - **Classifications**
    - **Other Aspects**
    - **Accounting Conventions**
    - **Scope of the Data**
    - **Nature of the Basic Data**
  - **Observance or Practice** (DC, CF)
  - **Plans** (DC, CF, Agency)

Methodological Source — **National Methodological publication** — **Cross Checks**

Terms and Conditions — **Prior Access** — **Ministerial Commentary** — **Notice of Revision**

Simultaneous Release — **ARC** — **Advance Release Information** — **Dissemination Formats**

Coverage — **Periodicity** — **Timeliness** — **Notes**

DSBB Data Model – Metadata Classes – Main Layers

Diagram B

Country/International Organization

Data

Metadata

Data (DC)

Access (DC, CF, Agency)

Integrity (DC, Agency)

Quality (DC)

Methodology (DC, CF)

Observance or Practice (DC, CF)

Plans (DC, CF, Agency)

Analytical Framework, Concepts and Definitions

Compilation Practices

Classifications

Other Aspects

Scope of the Data

Accounting Conventions

Nature of the Basic Data

Analytical Framework

Definitions

Concepts

Seasonal Adjustment

Verification and Consistency of Final Results

Revision Policy and Procedure

Geographical and Commodity Detail

Other Measures

Assessment and Validation of Intermediate Results

Estimation Methods

Aggregation Techniques

Statistical Adjustments to Basic Data Sources

Consolidation Methods

Unit Coverage

Transaction Coverage

Geographic Coverage

Unrecorded Activity

Industrial Coverage

ITRM (G&S) Coverage

Population Coverage

Topical Coverage

Occupational Coverage

Sectoral Coverage

Institutional Coverage

Instrument Coverage

Conversion Factors

Exceptions & Special Cases

Data Type (S/F)

Valuation Revaluation Method

Recording Conventions

Data Sources Conventions

Weighting

Sampling Methods
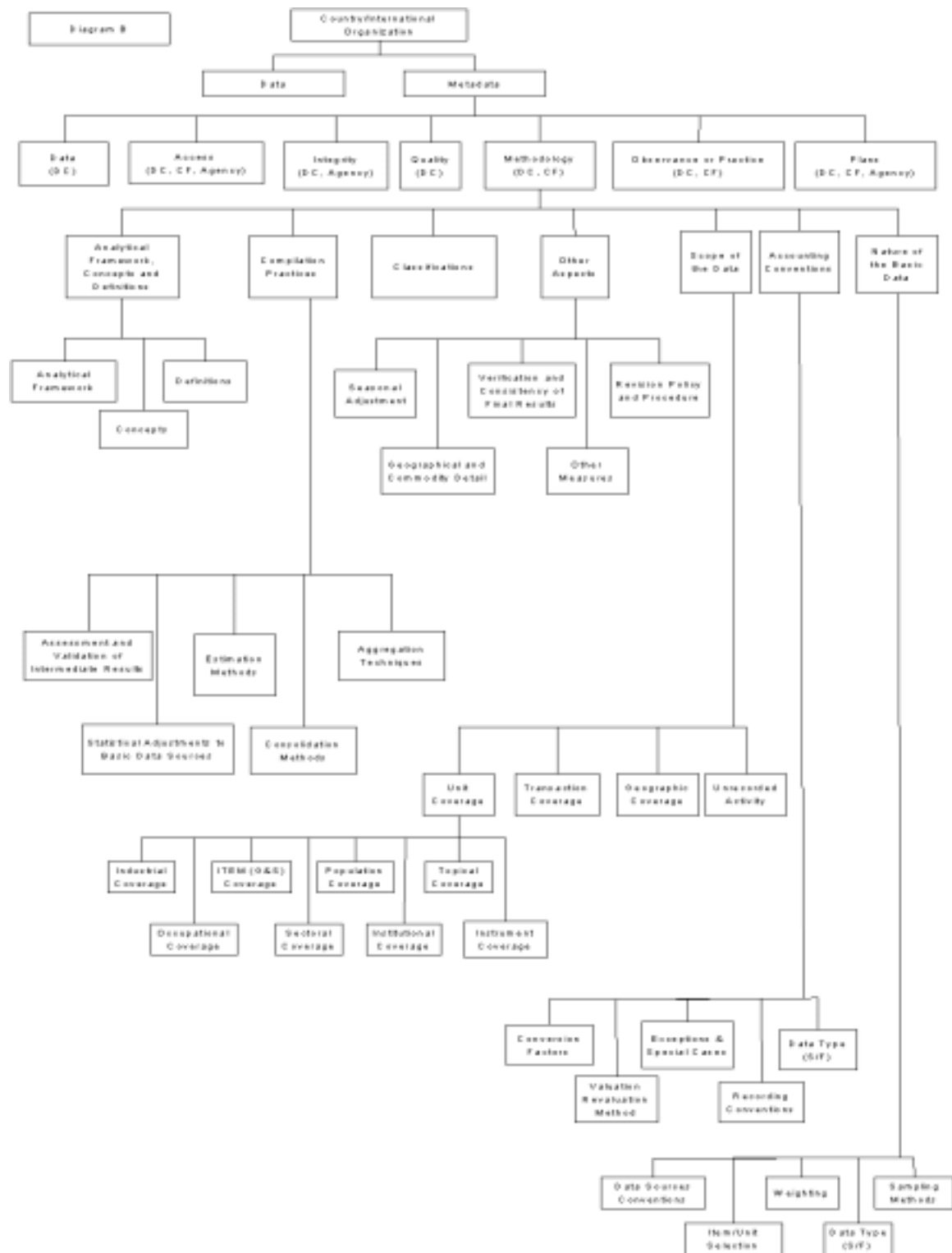
Item/Unit Selection

Data Type (S/F)

Table 1 (Work-in-progress)

Search and Query the DSBB - updated February 12, 2002

14.   It is now possible to search and query the metadata provided by SDDS subscribers and GDDS participants as well as by sector/data category and across sectors/data categories. A large number of pre-defined queries have been established based on frequently requested information. It is also possible to structure queries using the template titles appearing below. As the functionality of this facility is under continuous development, your comments and suggestions are welcome (DDSD@IMF.ORG)

```
SDDS Subscribers
GDDS Participants                         GO
SDDS Subscribers and GDDS Participtants
```

Sector
```
All
Real Sector
Fiscal Sector
```

Data Category
```
All
National Accounts
Industrial Production
```

Metadata Class
```
Coverage
Periodicity
Timeliness
```

Metadata Element
```
Base Year
Frequency of rebasing
Number of establishments surveyed
```

```
Search    Clear   Tips
```

Table 2 (Work-in-progress)

Search and Query the DSBB - updated February 12, 2002

15.   It is now possible to search and query the metadata provided by SDDS subscribers and GDDS participants as well as by sector/data category and across sectors/data categories. A large number of pre-defined queries have been established based on frequently requested information. It is also possible to structure queries using the template titles appearing below. As the functionality of this facility is under continuous development, your comments and suggestions are welcome (DDSD@IMF.ORG)

SDDS Subscribers	GO

Sector

| All |
| Real Sector |
| Fiscal Sector |

Data Category

| Wages and Earnings |
| Consumer Price index |
| Producer Price Index |

Metadata Class

| Coverage |
| Periodicity |
| Timeliness |

Metadata Element

| Base Year |
| Number of establishments surveyed |
| Number of commodities surveyed |

You have selected a query regarding information on the base year, number of establishments covered, and the number of commodities included in the Producer Price Indices of all 50 SDDS subscribers. To modify the existing search criteria, select from the options above and click on 'Search' below.

Search   Clear   Tips

Table 3 (Work-in-progress)

Results of a PPI query for SDDS Subscribers (partial results)

Special Data Dissemination Standard
Subscribers

SDDS Subscriber	Base year	Number of establishments surveyed	Number of commodities surveyed
Argentina	1993 = 100	1400	2800
Australia	1988-89 = 100	3200	1400
Austria	2000=100	270	373
Belgium	1980=100	280	1096
Brazil	August 1994=100	1651	462
Canada	1997=100	Approximately 2700	980
Chile	June 1992=100	685	378
Columbia	June 1999=100	2200	730
Costa Rica	December 1999=100	216	1000
Croatia	Previous year=100	320	400
Turkey	1994=100	1300	687
United Kingdom	1995=100	Approximately 3000	980
U.S.A.	1982=100	More than 25,000	Approximately 100,000

IMF Home | Search | Site Map | Site Index | Help | What's New

International Monetary Fund

About the IMF | News | Publications | Country Info | IMF Finances | Standards & Codes

Table 4 (Work-in-progress)

Modify Existing Search Criteria

Results of a PPI query for SDDS Subscribers (partial results)