



Committee for the Coordination of Statistical Activities
Conference on Data Quality for International Organizations
Wiesbaden, Germany, 27 and 28 May 2004



Session 3: Tools and policies for collecting, managing and disseminating data and metadata (2)

**DATA AND METADATA COLLECTION AND DISSEMINATION IN THE IMF STATISTICAL
INFORMATION SYSTEM**

THE IMF METADATA REPOSITORIES PROJECT:

**IMPROVING METADATA CONTENT AND ACCESS
THROUGH THE WORK OF THE SDMX INITIATIVE**

**Prepared by the Statistics Department
International Monetary Fund**

TABLE OF CONTENTS

Introduction	2
Background	2
The IMF’s SDDS/GDDS metadata repositories	2
Some characteristics of the SDDS and the DQAF	3
Main objectives of the IMF Metadata Repositories Project.....	4
Status of the project and way forward	5
Conclusions	6
Figure 1. The components of the SDMX deliverables.....	5
Figure 2. Class diagram of the Reference Metadata Set.....	8
Table 1. Example of an adaptation of the DQAF structure to the content of the SDDS.....	7
Appendix I	
Data Quality Assessment Framework—Generic Framework (July 2003 Framework).....	9
Appendix II	
The Cascading Structure of the Data Quality Assessment Framework, DQAF July 2003, for the National Accounts Statistics: An Example	12

Introduction

1. The Statistical Data and Metadata Exchange (SDMX) initiative brings together the Bank for International Settlements (BIS), the European Central Bank, the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the Statistical Office of the European Communities, the United Nations Statistics Division, and the World Bank (WB) with the objective of focusing on business practices in the field of statistical information that would allow more efficient processes for exchange and sharing of data and metadata within the current scope of their collective activities.
2. The goal is to explore common e-standards and ongoing standardization activities that could allow SDMX partners to gain efficiency and avoid duplication of effort in their own work and possibly for the work of others in the field of statistical information and its management. In the case of the metadata repositories project, SDMX partners intend to do this by taking advantage of existing metadata dissemination formats, such as that implicit in the IMF's Dissemination Standards Bulletin Board (DSBB).
3. This paper provides some background to the IMF metadata repositories project, which is one of the joint SDMX projects to address different aspects of statistical data and metadata exchange. It describes the existing metadata environment at the IMF, the work on the development of a model for the IMF's metadata repositories in support of open access to metadata, and how this work will contribute to enhancing the quality of metadata content and access.

Background

4. Information available on the SDMX website (www.SDMX.org) provides the most current information on the various projects taking place under the auspices of the SDMX initiative. In the course of 2004-2005, the SDMX members intend to deliver version 1.0 of SDMX standards and related tools and demonstrate their practical relevance in a pilot project implementation within the context of the joint BIS-IMF-OECD-WB statistics on external debt, which are a product of the Inter-Agency Task Force on Finance Statistics (TFFS). In the context of the work on the release of version 1.0 standards for the SDMX common information model, data formats, and core metadata, the IMF Metadata Repositories Project is referred to as an "SDMX alignment project". It will implement the proposed SDMX standards for the metadata covered by the IMF's Special Data Dissemination Standard (SDDS) and General Data Dissemination System (GDDS) in support of the open exchange of metadata. Until the initial version 1.0 SDMX standards are approved, inputs from the work on the IMF Metadata Repositories Project could be used as input for the joint external debt pilot project that will develop generic version 1.0 standards and implement them for that statistical domain.

The IMF's SDDS/GDDS metadata repositories

5. Subscription to the IMF's Special Data Dissemination Standard (SDDS) has grown to 57 countries, with The Kyrgyz Republic becoming the most recent subscriber on 26 February 2004. Metadata for all subscribing countries are on the Dissemination Standards Bulletin Board (DSBB), which was launched in September 1996. Further, metadata for 68 participants in the IMF's General Data Dissemination System (GDDS) built on the same four metadata aspects—data characteristics,

quality, access, and integrity—are posted on the DSBB. Taken together, this means that more than two-thirds of the IMF’s 184 members disseminate information about their statistical practices (or at least some of them) on the DSBB using the SDDS/GDDS framework. Moreover, the IMF is working with a number of additional countries that have committed to subscribing to the SDDS or using the GDDS as a framework for the development of their national statistical systems.

6. The use of a standard presentation format for statistical metadata on the DSBB enables data users worldwide to gain access to information in a readily recognizable and comparable form. In recognition of this, in March 2003 the IMF launched an enhanced DSBB website that transformed the existing set of static pages into pages that are dynamically generated on request. This has greatly increased the DSBB’s flexibility to meet specific user needs by enabling users to order up a set of web pages containing information gleaned from a search and query operation in formats tailored to their purpose.

7. Despite the success of these enhancements, the existing DSBB metadata query facility lacks the capability—and compatibility—to interact fully with other sources of statistical metadata available at the national and international levels. This is because a common vocabulary, an internationally agreed model articulating the manner in which information is stored, and a standard format for rendering metadata and macroeconomic time series data have not yet been sufficiently developed.

Some characteristics of the SDDS and the DQAF

8. Preliminary work on the implementation of SDMX standards for metadata exchange is done at the IMF, in collaboration with SDMX partners, taking advantage of the organization’s experience with metadata repositories. The IMF’s SDDS metadata repository covers eighteen data categories, encompassing a wide array of macroeconomic indicators. For most of these data categories, the SDDS is not prescriptive in terms of the dataset’s content, presentation (although minimal features of coverage are prescribed), recommended methodology, or compilation practices. These features point to the fact that a metadata model supporting the SDDS would be expected also to support a large number of datasets produced by national statistical agencies and central banks.

9. The IMF’s Data Quality Assessment Framework (DQAF) is an important consideration in producing a metadata model of the SDDS. The DQAF evolved after the SDDS and GDDS, and its main purpose is to provide a flexible structure for assessing data quality by comparing country statistical practices with best practices, including internationally accepted methodologies. Rooted in the United Nations *Fundamental Principles of Official Statistics*¹, it is the product of an intensive consultation with national and international statistical authorities and data users inside and outside the IMF. The DQAF has a cascading structure, moving from the dimensions common to all datasets, as captured in the Generic Framework (see Appendix I), to the more detailed aspects appropriate to

¹ The Fundamental Principles of Official Statistics were adopted by the United Nations Statistical Commission in a special session in 1994. The principles are intended to guide producers of official statistics in fulfilling their obligations and to inform users of statistics of what they should expect.

individual datasets in the dataset-specific DQAFs. Appendix II provides a view of the cascading structure of the DQAF.

10. The IMF has developed six dataset-specific DQAFs. These dataset-specific DQAFs cover national accounts, consumer price indices, producer price indices, government finance statistics, monetary statistics, and balance of payments. Another DQAF has been jointly prepared with the World Bank to cover consumer expenditures under poverty, and additional DQAFs are under development with other international organizations.

11. The cooperation at the international level in extending the DQAF as a framework for assessing data quality provides rationale for using the generic framework as a logical starting point for the elaboration of a metadata model for macroeconomic statistics and SDDS metadata. It is noteworthy that, while the SDDS predates the DQAF, all metadata elements of the SDDS represent subsets of metadata concepts of the DQAF. The metadata concepts of the DQAF are organized around a structure that is hierarchical and discrete, each item covering a particular aspect of data quality; SDDS metadata on the DSBB, on the other hand, are organized around key metadata features, without adhering to the logical structure of the DQAF. The model supporting SDDS metadata open exchange is expected to be largely derived from the structure of the DQAF, which would also support GDDS metadata exchange.

Main objectives of the IMF Metadata Repositories Project

12. The immediate objective of the metadata repositories project is twofold: (1) identify the commonalities in the metadata structures of macroeconomic datasets that are collected and stored in existing repositories; (2) use these commonalities to develop standardization in format, structure, and vocabulary. The long-term objective is to enable websites of national and international authorities, as well as private sector organizations, to support intelligent queries across diverse repositories using a common thesaurus of metadata terms. This would support the data exchange, publishing, and dissemination needs of a wide community of providers and users.

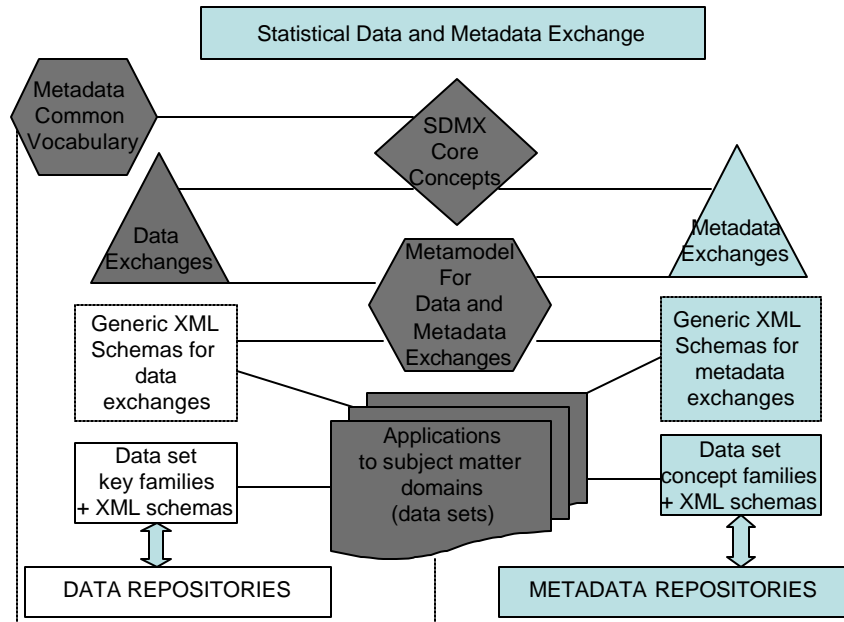
13. In its initial phase, the project will adhere to the all-encompassing approach adopted for SDDS metadata and by the generic DQAF; SDMX partners will work on developing a metadata model that could be extended to any dataset, without working on domain or dataset specific extensions of that model.² That future work will be coordinated by/with the secretariats of the international bodies responsible for overseeing the domain specific guidelines, such as the Inter-secretariat Working Group on National Accounts, the Inter-secretariat Working Group on Price Statistics, the Inter-Agency Task Force on Finance Statistics, the International Labour Office, the IMF Committee on Balance of Payments Statistics, and the Inter-agency Task Force on Statistics of International Trade in Services.

14. Figure 1 provides an illustration of the array of work covered by the various projects under the auspices of the SDMX initiative. The joint external debt hub pilot project is expected to deliver

² Extensions of the metadata model would apply to concepts such as “methodology” or “compilation practices”, for which the relevant component concepts are largely data category specific.

the standards required to support data exchanges (data repositories) as well as to implement these standards in a practical case study. As such, the activities of the pilot project cover most areas of Figure 1. The metadata repositories project will implement the SDMX standards in the SDDS metadata environment with the objective of providing open access to metadata repositories.

Figure 1. The components of the SDMX deliverables



Status of the project and way forward

15. The work on the SDDS metadata repository has led to a representation of the SDDS metadata framework that supports linkages to the DQAF structure at all levels of SDDS metadata. The next step involves using the structure of the DQAF to elaborate a metadata model that will support metadata exchanges for an as wide as possible range of statistical products. The preliminary work on the model has led to the identification of two broad classes of metadata for reference metadata—or, more specifically, for that category of reference metadata covered by the IMF metadata repositories.³ These are the “data category” (or dataset) concept and the “contact” concept, with the possibility of using “group keys” for metadata elements for which the content is identical across data categories.

16. The metadata content of each classes of reference metadata has not yet been fully determined and the relationships between the classes of metadata have to be further worked out. In

³ To facilitate understanding the scope of its activities, the SDMX initiative has tentatively come up with the term “reference metadata” to refer to metadata that provides information on data access, compilation methodologies, and agencies assuming responsibility for the production of the datasets. The scope of reference metadata is very large, as these metadata cover all statistical information not directly associated with those supporting data exchanges.

proposing concepts for metadata classes and content, the IMF is relying on the international consensus that exists around the SDDS as a standard for metadata dissemination and the DQAF as a tool to assess data quality. Throughout the development of the model and its practical application to SDDS metadata, SDMX partners will work closely together and with national statistical offices and central banks. Much use will be made of the www.SDMX.org website to communicate the latest developments in the project and solicit the collaboration of the statistical community.

17. Table 1 provides an example of SDDS metadata expressed in terms of the DQAF structure. This formulation of a generic reference metadata model is very preliminary and provided for illustration purposes. It immediately brings forward the need for the Metadata Common Vocabulary (MCV) to include the various concepts included in Table 1—or for these concepts to be replaced by terms defined in the MCV. This example does not show the links between the classes of reference metadata, nor the attributes attached to the various metadata concepts. Figure 2 provides an example of a UML (Unified modeling language) illustration of associations and attributes that could be attached to various metadata concepts when building a reference metadata model.

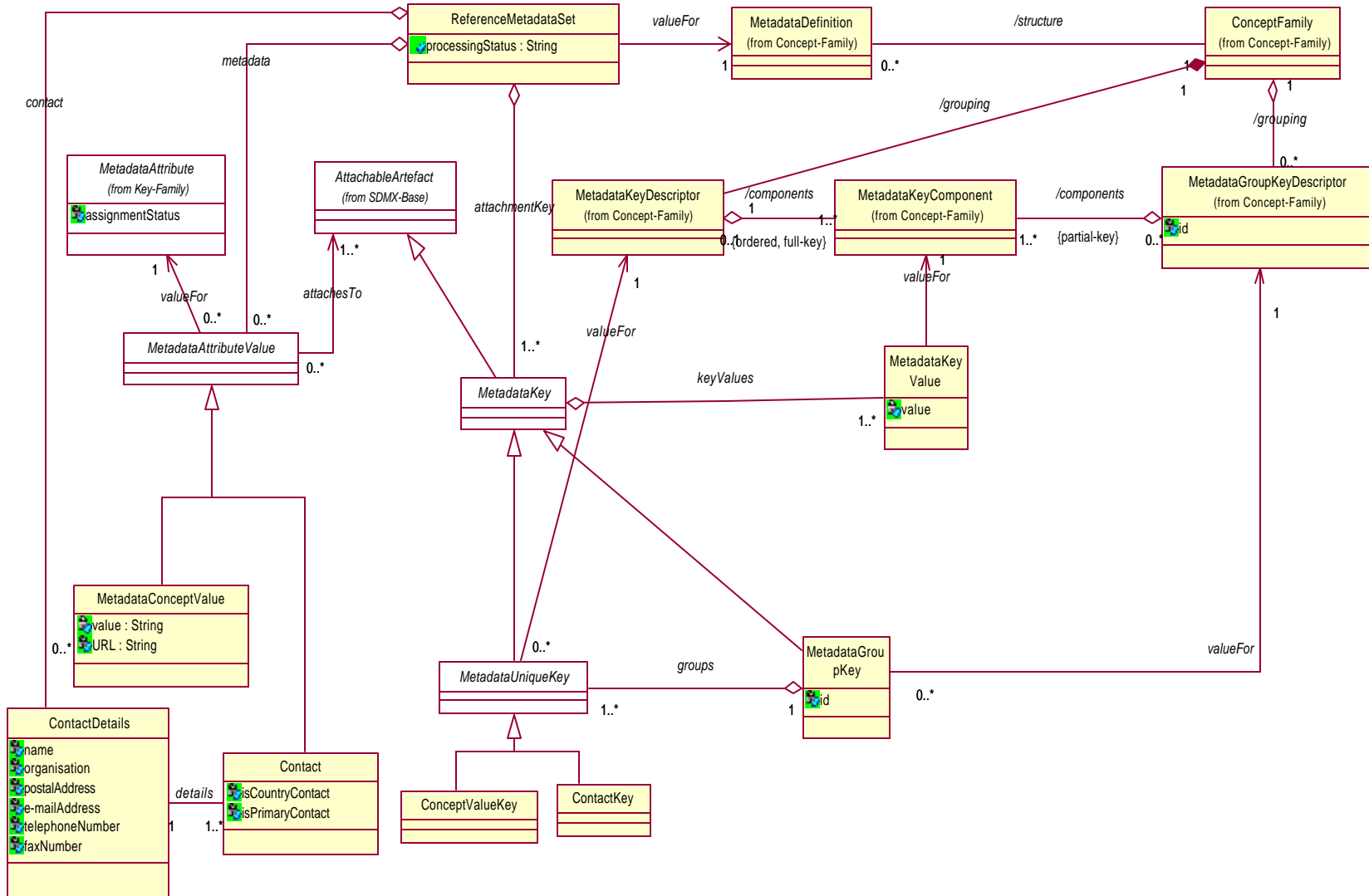
Conclusions

18. The work on developing a metamodel for reference metadata, while solidly anchored in internationally accepted standards for metadata dissemination practices, is still in its early phase. The work presented in this paper is subject to change and refinement, but SDMX partners want to inform the statistical community of the direction of their work at an early stage and solicit the views and collaboration of all interested parties.

Table 1. Example of an adaptation of the DQAF structure to the content of the SDDS

Metadata Concept
Data category concept
Agency
Legal environment
Terms and conditions
Confidentiality
Mandatory reporting
Integrity
Transparency
Internal government access
Ministerial commentary
Serviceability
Periodicity
Timeliness
Revision policy
Accessibility
Data
Presentation
Advance release calendar
Simultaneous release
Metadata
Assistance
Publications
Methodology
Concepts
Scope
Classification
Basis for recording
Accuracy/Reliability
Source data
Statistical techniques
Assessment and validation
Contact concept
Name
Address
e-mail

Figure 2. Class diagram of the Reference Metadata Set



**Data Quality Assessment Framework—Generic Framework
(July 2003 Framework)**

Quality Dimensions	Elements	Indicators
<p>0. Prerequisites of quality</p>	<p>0.1 Legal and institutional environment—<i>The environment is supportive of statistics</i></p> <p>0.2 Resources—<i>Resources are commensurate with needs of statistical programs.</i></p> <p>0.3 Relevance—<i>Statistics cover relevant information on the subject field.</i></p> <p>0.4 Other quality management—<i>Quality is a cornerstone of statistical work.</i></p>	<p>0.1.1 The responsibility for collecting, processing, and disseminating the statistics is clearly specified.</p> <p>0.1.2 Data sharing and coordination among data-producing agencies are adequate.</p> <p>0.1.3 Individual reporters' data are to be kept confidential and used for statistical purposes only.</p> <p>0.1.4 Statistical reporting is ensured through legal mandate and/or measures to encourage response.</p> <p>0.2.1 Staff, facilities, computing resources, and financing are commensurate with statistical programs.</p> <p>0.2.2 Measures to ensure efficient use of resources are implemented.</p> <p>0.3.1 The relevance and practical utility of existing statistics in meeting users' needs are monitored.</p> <p>0.4.1 Processes are in place to focus on quality.</p> <p>0.4.2 Processes are in place to monitor the quality of the statistical program.</p> <p>0.4.3 Processes are in place to deal with quality considerations in planning the statistical program.</p>
<p>1. Assurances of integrity</p> <p><i>The principle of objectivity in the collection, processing, and dissemination of statistics is firmly adhered to.</i></p>	<p>1.1 Professionalism—<i>Statistical policies and practices are guided by professional principles.</i></p> <p>1.2 Transparency—<i>Statistical policies and practices are transparent.</i></p> <p>1.3 Ethical standards—<i>Policies and practices are guided by ethical standards.</i></p>	<p>1.1.1 Statistics are produced on an impartial basis.</p> <p>1.1.2 Choices of sources and statistical techniques as well as decisions about dissemination are informed solely by statistical considerations.</p> <p>1.1.3 The appropriate statistical entity is entitled to comment on erroneous interpretation and misuse of statistics.</p> <p>1.2.1 The terms and conditions under which statistics are collected, processed, and disseminated are available to the public.</p> <p>1.2.2 Internal governmental access to statistics prior to their release is publicly identified.</p> <p>1.2.3 Products of statistical agencies/units are clearly identified as such.</p> <p>1.2.4 Advanced notice is given of major changes in methodology, source data, and statistical techniques.</p> <p>1.3.1 Guidelines for staff behavior are in place and are well known to the staff.</p>

Quality Dimensions	Elements	Indicators
<p>2. Methodological soundness</p> <p><i>The methodological basis for the statistics follows internationally accepted standards, guidelines, or good practices.</i></p>	<p>2.1 Concepts and definitions—<i>Concepts and definitions used are in accord with internationally accepted statistical frameworks.</i></p> <p>2.2 Scope—<i>The scope is in accord with internationally accepted standards, guidelines, or good practices.</i></p> <p>2.3 Classification/sectorization—<i>Classification and sectorization systems are in accord with internationally accepted standards, guidelines, or good practices.</i></p> <p>2.4 Basis for recording—<i>Flows and stocks are valued and recorded according to internationally accepted standards, guidelines, or good practices</i></p>	<p>2.1.1 The overall structure in terms of concepts and definitions follows internationally accepted standards, guidelines, or good practices.</p> <p>2.2.1 The scope is broadly consistent with internationally accepted standards, guidelines, or good practices.</p> <p>2.3.1 Classification/sectorization systems used are broadly consistent with internationally accepted standards, guidelines, or good practices.</p> <p>2.4.1 Market prices are used to value flows and stocks.</p> <p>2.4.2 Recording is done on an accrual basis.</p> <p>2.4.3 Grossing/netting procedures are broadly consistent with internationally accepted standards, guidelines, or good practices.</p>
<p>3. Accuracy and reliability</p> <p><i>Source data and statistical techniques are sound and statistical outputs sufficiently portray reality</i></p>	<p>3.1 Source data – <i>Source data available provide an adequate basis to compile statistics.</i></p> <p>3.2 Assessment of source data—<i>Source data are regularly assessed.</i></p> <p>3.3 Statistical techniques—<i>Statistical techniques employed conform to sound statistical procedures</i></p> <p>3.4 Assessment and validation of intermediate data and statistical outputs—<i>Intermediate results and statistical outputs are regularly assessed and validated.</i></p> <p>3.5 Revision studies—<i>Revisions, as a gauge of reliability, are tracked and mined for the information they may provide.</i></p>	<p>3.1.1 Source data are obtained from comprehensive data collection programs that take into account country-specific conditions.</p> <p>3.1.2 Source data reasonably approximate the definitions, scope, classifications, valuation, and time of recording required.</p> <p>3.1.3 Source data are timely.</p> <p>3.2.1 Source data—including censuses, sample surveys, and administrative records—are routinely assessed, e.g., for coverage, sample error, response error, and nonsampling error; the results of the assessments are monitored and made available to guide statistical processes.</p> <p>3.3.1 Data compilation employs sound statistical techniques to deal with data sources.</p> <p>3.3.2 Other statistical procedures (e.g., data adjustments and transformations, and statistical analysis) employ sound statistical techniques.</p> <p>3.4.1 Intermediate results are validated against other information where applicable.</p> <p>3.4.2 Statistical discrepancies in intermediate data are assessed and investigated.</p> <p>3.4.3 Statistical discrepancies and other potential indicators or problems in statistical outputs are investigated.</p> <p>3.5.1 Studies and analyses of revisions are carried out routinely and used internally to inform statistical processes (see also 4.3.3).</p>

Quality Dimensions	Elements	Indicators
<p>4. Serviceability <i>Statistics, with adequate periodicity and timeliness, are consistent and follow a predictable revisions policy.</i></p>	<p>4.1 Periodicity and timeliness— <i>Periodicity and timeliness follow internationally accepted dissemination standards.</i></p> <p>4.2 Consistency— <i>Statistics are consistent within the dataset, over time, and with major datasets.</i></p> <p>4.3 Revision policy and practice— <i>Data revisions follow a regular and publicized procedure.</i></p>	<p>4.1.1 Periodicity follows dissemination standards. 4.1.2 Timeliness follows dissemination standards.</p> <p>4.2.1 Statistics are consistent within the dataset. 4.2.2 Statistics are consistent or reconcilable over a reasonable period of time. 4.2.3 Statistics are consistent or reconcilable with those obtained through other data sources and/or statistical frameworks.</p> <p>4.3.1 Revisions follow a regular and transparent schedule. 4.3.2 Preliminary and/or revised data are clearly identified. 4.3.3 Studies and analyses of revisions are made public (see also 3.5.1).</p>
<p>5. Accessibility <i>Data and metadata are easily available and assistance to users is adequate.</i></p>	<p>5.1 Data accessibility— <i>Statistics are presented in a clear and understandable manner, forms of dissemination are adequate, and statistics are made available on an impartial basis.</i></p> <p>5.2 Metadata accessibility— <i>Up-to-date and pertinent metadata are made available.</i></p> <p>5.3 Assistance to users— <i>Prompt and knowledgeable support service is available.</i></p>	<p>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 format are adequate. 5.1.3 Statistics are released on a pre announced schedule. 5.1.4 Statistics are made available to all users at the same time. 5.1.5 Statistics not routinely disseminated are made available upon request.</p> <p>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.</p> <p>5.3.1 Contact points for each subject field are publicized. 5.3.2 Catalogs of publications, documents, and other services, including information on any changes, are widely available.</p>

**The Cascading Structure of the Data Quality Assessment Framework,
DQAF July 2003, for the National Accounts Statistics: An Example**

Using serviceability as the example of a dimension of quality, the box below shows how the framework identifies three elements that point toward quality. Within consistency, one of those elements, the framework next identifies three indicators. Specifically, for each indicator, focal issues are addressed through key points that may be considered in identifying quality.

