



**Economic and Social
Council**

Distr.
GENERAL

CES/2002/8
12 April 2002

ORIGINAL: ENGLISH

**STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

Fiftieth plenary session
(Paris, 10-12 June 2002)

**PROGRESS REPORT ON NEW DEVELOPMENTS IN DATA AND METADATA
COLLECTION FOR INTERNATIONAL ORGANISATIONS**

Submitted by the IMF and OECD¹

I. Introduction

1. Statistical data and the associated metadata are essential resources for the analytical work of International Organisations and their collection represents a significant part of the regular statistical work of those Organisations. Providing data to International Organisations represents a significant task for national data providers². Countries are sometimes expected to deal with multiple requests for the similar data from different International Organisations and, in some cases, from different areas within the same Organisation. It is an objective of International Organisations working in the area of statistics to collectively minimise the data reporting burden on national providers. This paper is an attempt at identifying the main issues in establishing joint data reporting arrangements and presents some possible directions for immediate and medium term action in the area of short-term statistics.

II. Current organisation of data and metadata collection at the IMF, OECD, and within the community of International organisations

II.1 A complex system of International Organisations and data suppliers

2. The system of International Organisations has been developed by the international community of governments to serve various political needs. Their structures vary enormously, as do their mandates and the position of statistical activities in their mandate, their legal power and, finally, the resources available for their statistical activities. In this paper the International Organisations covered are: the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organisation for Economic

Co-operation and Development (OECD), the United Nations System of Organisations, and the World Bank (WB). The paper is mainly based on the experiences of the IMF and OECD.

3. The internal structure of individual International Organisations adds to the complexity of the system. For example, the OECD Secretariat has a decentralised statistical structure similar to that of central administrations in many of its Member countries. Statistical activities at the OECD are decentralised, statisticians and analysts working closely together on a subject by subject basis to support policy analysis discussions conducted by Committees of policy makers or senior civil servants representing each OECD Member Country. Data are collected from Member countries under the authority of those committees. There is no Committee working on statistics as a subject in its own right, but there is a Statistics Directorate and a Chief Statistician whose role is to co-ordinate OECD statistical activities. A Statistical Policy Group, composed of senior statisticians from subject matter areas, helps the OECD Chief Statistician on his internal statistical co-ordination tasks. The OECD Statistical Advisory Group (SAG) and the High-level OECD statistical meetings are additional instruments for the development co-operation between the OECD and national statistical authorities.

4. The IMF is organized along geographical and functional lines. Each of these areas carries out its own data collection activities according to its own operational requirements. The geographical departments monitor developments in country economies and collect data that are highly time-sensitive. Data exchanges are worked out bilaterally with each country and institution, and may be supplemented by commercial data sources. The Statistics Department has functional responsibilities, i.e., dealing with statistics. It collects data of record primarily for the purpose of publication. The magnitude and purpose of its data collection exercises necessitate a more uniform system, although some bilateral arrangements are in place. The Statistics Department has pursued joint data collection with other IMF departments, where feasible, e.g., where time sensitivity is not of paramount importance.

5. In addition to different internal structures and mandates for collecting specific data, each organization has its preferred technology tools, nomenclature, data structures, etc.. From a data provider's point of view, the proliferation of systems, reporting formats and data requests can be overwhelming. Data providers are not a homogeneous category either. They include NSOs but also central banks, and many ministries. Responsibilities regarding subject matters differ from country to country, as do the organizational structure of the national agencies and their statistical systems.

II.2 Overlapping data requirements

6. There are large areas of overlap between the data requirements of different organisations. For example, the OECD *Main Economic Indicators* and the IMF *International Financial Statistics* publications share a large number of indicators, e.g. Balance of Payments, Consumer Prices, Monetary Aggregates, Industrial Production, International Trade Totals and Labour Force. The same comparison could be made, for "main economic indicators" in particular, between the data requirements of Eurostat, the ECB and the OECD. Another example of data needed by most international organisations is National Accounts data. There are also important differences, for example, country coverage for the IMF and the UN is more extensive than that of the OECD and Eurostat.

II.3 A variety of requirements and systems

7. Although national sources may provide very similar data to several organizations, there are also a number of differences that make each data request unique. The details and processes of data collection of different International Organisations vary considerably in response to varying needs. The nature of the details and processes involved depends on a number of criteria:

- nature of data to be collected, e.g. whether or not they are part of a regular production process;
- frequency of the collection, which varies from daily to annual or even occasional;
- regularity of the content over time;
- homogeneity of content regarding the internal structure of the NSO (e.g. subject matters) ;
- quantity and level of data (e.g., disaggregated versus aggregated data);
- method of collection;
- security and confidentiality requirements;
- technology and infrastructure.

8. In addition to the diversity generated by those criteria, processes for data collection have evolved over the years as the result of bilateral arrangements between individual organisations and national agencies or other international organisations. Of course, progress in information technology and communication has been an important driver of those evolutions. Also, policies regarding arrangements vary from accepting any format and medium suggested by the data provider to legally imposing a fixed format and fixed medium.

9. The OECD, for example, has been very flexible in the area of short-term statistics and has preferred to adjust to formats readily available to individual data providers in order to gain on timeliness. In some cases, the IMF directly accesses national and/or transnational databases in order to obtain the statistics it needs for its operational activities, although this approach requires more resources and could not be sustained if applied across its membership. On the other hand, regional and/or transnational agencies are more likely to prescribe fixed formats and medium. Eurostat, for example, recommends the use of a precise format, in general one of the implementations of GESMES³. Eurostat also recommends the use of its data collection infrastructure, in particular the Stadium system which is a centralised service for collecting and administering statistical data files submitted to Eurostat. Stadium also redistributes incoming data files to the appropriate Eurostat areas. However, not all EU countries follow Eurostat's recommendations. The ECB, on the other extreme, imposes that all data reporting use GESMES/CB.

10. Despite many efforts to develop standards and to increase co-operation, the international statistical community currently uses an enormous number of data streams, formats and technologies to exchange statistical data and metadata. Media include paper, diskettes, CD-ROMs, mainframe magnetic tapes, email, other file transfer methods, web databases and other on-line databases. Formats vary from text with separated or fixed fields (e.g. CSV), proprietary formats such as MS-Excel or Fame databases, on-line questionnaires and of course the statistical standard GESMES in its various versions. The structure of the data received vary from tables,

sets of times series and predefined questionnaires.

II.4 Metadata

11. Metadata are in general not as well structured as data. They can refer to a whole data set, a dimension, a value of a dimension, a projection of the whole file on one or several dimensions, or finally they can refer to one or several individual data points. Given the complex nature of metadata their collection is a much more complicated task. In printed publications, data and metadata are in general associated. Over the years, with the generalisation of electronic files for the submission of data, data and metadata have been dissociated. It is only recently, with the emergence of new formats for statistics such as GESMES and of new software for statistical data, that they can be provided together in electronic files. Another important element in the area of metadata, which is knowledge about data, is the information about contact people. With the move of data collection towards electronic means the main actors of the data exchange have become technicians and the contact between the people who know about the statistical content has been loosened.

12. A remarkable initiative in the recent past in the area of metadata is the IMF's Special Data Dissemination Standard⁴ (SDDS) and the General Data Dissemination System (GDDS). In this system, countries provide metadata to the IMF about their dissemination practices using a set of standard formats. The IMF presents these metadata using a common web page layout for each of the SDDS data categories. These web pages are often linked to country web sites where the data described by the metadata may be viewed. However, the SDDS does not cover all the economic variables needed by other International Organisations such as the OECD and Eurostat. Furthermore, the metadata are a special subset designed to focus on dissemination practices, hence they would not necessarily be complete for data collection purposes.

II.5 Current practices to monitor data flows and the quality of data capture procedures

13. International Organisations currently have their own methods for monitoring data flows. With the Ediflow project, Eurostat is using a database system in which each occurrence of data exchange is registered. Using Ediflow Eurostat can report to NSOs on the problems and on the quality of data transfers.

14. At the OECD, information on data collection has been collected for all OECD statistical activities in the context of the OECD Statistical Work Programme database⁵. Information collected can be used to investigate possible duplication in data collection by different parts of the organisation from the same source and to identify the range of collection instruments used.

15. In addition, for OECD short-term statistics, each individual data exchange arrangement is described in a database, with, inter alia, the exact technical documentation, contact names for technical and content information. The development of this database at the OECD has been done using some elements of the Ediflow system. Given the fact that the objective at the OECD was not to register all data flows, mainly because of resource and structure constraints, it was not possible to use exactly the same system as used by Eurostat.

16. The IMF Statistics department monitors data flows within each topical area and shares

(rather than duplicates) commonly used indicators across topical areas. There are isolated instances of joint data collection across IMF departments; they are managed by the Statistics Department and occur mostly in the area of financial statistics. There is no overall picture of data flows for all data collection activities undertaken in the IMF. Although this information would be very useful to have, gathering it would be very labour intensive; the Statistics Department alone collects data from 180 countries, from 3 or more agencies within each country, and from multiple units within each agency.

17. The databases that do exist are very useful for co-ordinating data collection activities within Organisations, and between one Organisation and its data providers. However, in their present state, they cannot provide a complete picture of all data flows from NSOs, and other data sources, to International Organisations. A possibility for improving global co-ordination would be to have a system for providing feed back from International Organisations to data providers on the quality of their data flows. This would have the advantage for data providers to monitor the situation with several International Organisations and for International Organisations to have the attention of management in NSOs for data transfer problems, which could generate quality improvements.

III. Reducing the reporting burden and making data collection more efficient

III.1 Co-operation between International Organisations on the content

18. It is an on-going effort of the community of International Organisations to improve co-ordination of data collections. An important number of collections are already co-ordinated between International Organisations. Several bodies with representatives from national statistical authorities and international organisations are working on improving international co-ordination in the area of statistics. Those bodies include the UN ACC Subcommittee on Statistical Activities, the UN Statistical Commission for Europe, and the Conference of European Statisticians with the Integrated Presentation of International Statistical Work Programmes. Table 1, below, shows the list of OECD data collections that are co-ordinated with other international organisations⁶. Table 2 provides a similar table on IMF coordination activities.

19. Co-ordination may take several forms, in particular:

- co-ordination of definitions without co-ordinating the actual exchange of data;
- definition and implementation of common formats, the GESMES and more recently the SDMX initiatives;
- usage of common questionnaires and NSOs sending identical data files to different International Organisations – for example: annual national accounts, see the details below;
- one Organisation collecting and processing data from NSOs on behalf of other International Organisations - for example: co-ordination in the collection of annual international trade data between UNSD and the OECD;
- data exchange between international organisations, for example the OECD use the IMF's exchange rates data from the IMF's publication International Financial Statistics.

Table 1
OECD data collection activities co-ordinated with other International organisations

Statistical Activity	Co-ordinated with	Common quest.	Freq.
Annual Transport Statistics	UNECE	Yes	annually
	EUROSTAT	Yes	annually
Investment in Transport Infrastructure	EU	No	annually
	EIB	No	annually
Foreign Direct Investment	IMF	Yes	annually
	EUROSTAT	Yes	annually
Privatisation	WB	No	annually
Revenue Statistics	IMF	No	annually
Creditor Reporting System (CRS) Aid Activity Database official development assistance and official aid to developing countries and countries in transition]	WB	Yes	quarterly
External Debt of the Developing and Transition Countries, Transmission of World Bank/BIS/IMF data to OECD for publication in the Joint Statistics on External Debt and integration into OECD's compilations of external debt totals	WB	No	quarterly
	IMF	No	quarterly
Education UOE (UNESCO/OECD/EUROSTAT) Data Collection on Education Systems	UNESCO	Yes	annually
	EUROSTAT	Yes	annually
Health	WHO	No	annually
	UNECE	No	annually
	EUROSTAT	No	annually
Social Expenditures	EUROSTAT	No	annually
Annual Labour Force Statistics	ILO	No	annually
	EUROSTAT	No	annually
Annual National Accounts	UNSD	Yes	real-time
	EUROSTAT	Yes	real-time
International Trade by Commodity	UNSD	No	annually
	EUROSTAT	No	annually
Main Economic Indicators	ECB	No	monthly
	EU	No	monthly
	ILO	No	monthly
	IMF	No	monthly
	EUROSTAT	No	monthly
Main Economic Indicators for Non-Member Countries	EUROSTAT	No	monthly
Purchasing Power Parities	EUROSTAT	No	real-time
Quarterly Labour Force Statistics	EUROSTAT	No	monthly
Quarterly National Accounts	EUROSTAT	Yes	real-time
Statistics on Enterprises By Size Class	EUROSTAT	No	annually
Structural Statistics for Industry and Services	UNIDO	No	annually
	EUROSTAT	No	annually
Value Added and Employment in Services, done jointly with Annual National Accounts and STAN	EUROSTAT	Yes	annually
Territorial Statistics and Indicators	EUROSTAT	No	annually

Table 2
IMF data collection activities co-ordinated with other International and Transnational organizations

Statistical Activity	Co-ordinated with	Common quest. I/	Freq.
External Debt	BIS	Partial I/	Quarterly
Financial statistics	Regional central banks (e.g., BCEAO, BEAC, ECB, ECCB)	Yes, within regions, but not across regions	Monthly
International Reserves	ECB, ECCB, WAEMU (in progress)	Partial I/	Monthly
Labor Statistics	ILO	Yes	Quarterly
National Accounts (in progress)	EUROSTAT	Partial I/	Quarterly
Production and Prices (in progress)	OECD	No	Monthly
Population	UNSD	Yes	Annual
Trade Statistics	EUROSTAT	Yes	Monthly

I/ Denotes agreement in the data collected for a given topic, but not necessarily in the form of a joint questionnaire; “Partial” indicates that the IMF must calculate aggregates from reported items or that some items must still be collected from national sources.

20. In the case of the IMF, many of its coordination activities were time consuming to implement, partly because they were developed on a case-by-case basis. Each arrangement is slightly different, reflecting the variety of requirements and systems alluded to in section II.3. The lack of an established scheme for reaching agreements on the process and details has been a significant constraint on the IMF’s ability to pursue more agreements.

21. As described in section II, there is no mechanism for monitoring data flows between national data providers and International Organisations. Consequently, it is difficult to measure the burden on data providers and its evolution over time.

22. The following paragraphs describe four particular examples of co-ordination between International Organisations in the area of data collection:

- Annual International Trade

23. The collection of Annual International Trade statistics is co-ordinated between the OECD and the UN Statistics Division. The OECD collects from its Member Countries and forwards the data to the UN. The next objective for the UN and the OECD, in this area, is to have their databases and publications harmonised for OECD countries. This requires agreeing on definitions and processing techniques for, inter alia, inclusions/exclusions, treatment of

confidential data adjustments of quantity data, estimations of missing quantities, etc.

24. The IMF obtains international trade data from EUROSTAT for most EU countries. The lack of homogeneity among trading partners can create processing issues, although these are relatively minor in comparison to the benefits to the IMF of obtaining timely data.

- National Accounts

25. In the area of data collection for National Accounts, Eurostat, the OECD and the United Nations Statistics Division (UNSD) have co-ordinated their efforts. The tables requested are strictly identical. National Accounts files received by the Eurostat's Stadium system are automatically forwarded to the OECD. National Accounts data are sent by NSOs of EU countries and EU accession countries to Eurostat when particular accounts become available. The format used is GESMES. This creates a continuous flow of data throughout the calendar year. Once data are verified and ready for publication at the OECD, a copy of the OECD database is sent to UNSD. The format used is the native MS SQL Server format since both the UN and the OECD use that database system. In theory, this process was supposed to avoid having NSOs send files to the three organisations.

26. In practice, however, the transfer of National Accounts data from EU countries to the OECD via Eurostat does not always work well. There have been several instances where countries have sent data to Eurostat without the OECD receiving them. One reason is that, in some instances, countries have used other channels than Stadium and, consequently, the OECD has not received the data. This is particular frequent for Financial Accounts which are not always sent by NSOs but by Central Banks instead. Another problem with this theoretical automatic system is that the contact between the national agency and the OECD for National Accounts is loosened.

27. In order to solve those problems, the OECD would like EU member countries to send, via e-mail, the same data, the same day to Eurostat and to the OECD. This does not add any cost to member countries. Another alternative, would be that agencies place a copy of their National Accounts tables on a web site where the OECD, Eurostat and any other authorised International Organisation could get the data. The web site can simply be their standard web site. This second solution has the advantages of having the National agencies doing only one data transfer and of providing a central point where all tables are available for one particular country. If adopted for National Accounts it could be generalised to other statistical domains.

28. In the area of National Accounts a remarkable feature is the standard codification scheme used for data items. This codification makes the identification of data items very simple. It permits also an easy location of the related metadata.

29. The IMF is working with Eurostat to collect national account statistics for candidate countries. An important difference exists in the level at which each organization collects data, with Eurostat collecting highly disaggregated data and the IMF collecting only a handful of highly aggregated data. In order for a data exchange to work, the IMF must accept the disaggregated data from Eurostat and calculate the aggregates it needs. The IMF is also investigating possibilities for obtaining national accounts data from the OECD.

- Purchasing Power Parities

30. Under the Joint OECD-Eurostat PPP Programme, the OECD and Eurostat share the responsibility for collecting data for the calculation of PPPs. Broadly, Eurostat handles the calculations for the EU countries and for the EU Candidate Countries (i.e. those countries which have applied for admission to the EU). The OECD deals with the non-European OECD Member countries and the other non-EU related countries such as Russia, China, Ukraine, etc. which are included in the PPP Programme.

- Financial Statistics

31. The IMF has a number of data exchange agreements with regional central banks, such as the Banque Centrale des Etats de l'Afrique de l'Ouest (BCEAO), the Banque Centrale des États de l'Afrique Centrale (BEAC), the Eastern Caribbean Central Bank (ECCB) and the ECB. These regional banks provide data to the IMF on behalf of their members. These arrangements are based on an agreed set of standard items to be reported for all countries within the region. The ECB provides its data in GESMES/CB format over a VPN that it installed at the IMF for this purpose. The BCEAO, BEAC and ECCB provide data in the form of Excel spreadsheets. The system is very efficient in supporting automated data exchanges that reduce manual data processing for the IMF and retain the integrity of the figures as provided.

32. While joint data collection may alleviate the reporting burden on national sources by reducing the number of data requests, it can add to the data processing work of International Organizations at the receiving end, and in some cases may not obviate entirely the need to collect directly from national sources.

III.2 Some Lessons Learned

33. A key problem area for the IMF has been the high level of effort involved in setting up each data sharing arrangement. Agreements must be reached in a number of areas, including those listed in Section II.3. Some agreements require that one partner change or add to existing computer processes, these must also be agreed and documented. The work is made more complex by the differing internal structures of international organizations, which sometimes require multiple agreements within a single organization. There are no doubt many standard elements that make up each agreement; these elements could be elaborated, and possibly codified into a generic template that could be used to streamline the process of establishing data sharing agreements. The template could be supplemented by recommendations and/or "best practices", as relevant.

34. The template would need to cover at least the following general areas:

- CONTENT -- mapping the data sets for both partners to determine the overlap can be a very time consuming task; it often requires an item by item examination to identify differences; the problem is further exacerbated when partners collect data at different levels. Is there a more

efficient method of mapping, e.g., referencing a standard to which each topical area could map their data sets, for example, the SNA93 or Balance of Payments, Fifth Edition?

- NOMENCLATURE --part of the difficulty of mapping concepts is that each organization uses its own nomenclature to identify time series and to name critical metadata elements such as magnitude, scale, description, etc.; a key question is whose nomenclature to use in setting up the transfer—the reporters or the collectors? A scalable solution is needed, e.g., that an international organization could apply to 200+ reporters.
- FREQUENCY OF DATA AND/OR COLLECTION CYCLE – differences in the timing of updates must be agreed, as well as the frequency of the reported observations (monthly, quarterly, annual); differences in frequency can lead to further processing by the recipient, e.g., converting from a higher to a lower frequency, if low frequency data are needed.
- METHODOLOGY – many organizations add value to the data they collect by transforming it in some way. For example, data collected by the IMF for the *International Financial Statistics* publication are restated in common currencies, scaled to a common factor, rebased to a common base year, etc.. Some of these transformations may be aided in the data collection process itself (e.g., by metadata) and need to be retained in any data sharing arrangements. Similarly, some organizations produce very long historical time series and need historical revisions as far back as a country could provide. These and other methodological issues must be fully accounted for in joint data collection.
- CHANGE MANAGEMENT –changes to how items are compiled occur over time and a method must be found of informing the recipient of the necessary changes. In some instances changes are missed altogether, e.g., where only a standard set of items is exchanged, even though new items (often country specific) may be issued by the national source.
- TECHNOLOGY – agreements on the formats and medium to use will vary according to an international or transnational organizations capabilities. Ideally there would be agreement on a few formats, geared to different levels of technological capacity, from Excel spreadsheets (at the low end) to GESMES-based messages such as BOPSTA and GESMES/CB (at the high end).

III.3 Co-ordination between International Organisations on tools and formats

III.3.1 The GESMES message

35. In the early 1990s, the syntax for an EDIFACT message called Generic Statistical Message (GESMES) was developed under the leadership of Eurostat and with the participation of other International organisations, European NSOs and Central Banks. The IMF began using a profile of GESMES called BOPSTA for the collection of Balance of Payments, beginning in 1994-95. In 1998-99 another profile called GESMES/CB⁷ was introduced by the Bank for International Settlements, the European Central Bank and Eurostat. The IMF and OECD also use GESMES/CB for sending and receiving data.

36. The GESMES initiative is aimed at standardising formats for the exchange of statistical data. It reduces the burden on national sources in the sense that it reduces the number of formats requested. However, the main goal of GESMES is to make more efficient the treatment by receiving agencies and to make it easier to share data files between International Organisations.

37. Of notable interest here is the data model underlying the GESMES/CB (EDIFACT) message for the batch exchange of time series. The data model provides a very useful framework for exchanging time series, particularly between a national source and a transnational or international organization. A number of the elements that are needed to properly exchange data have been accounted for in the GESMES/CB data model. The user guide provides some recommendations to facilitate the adoption of GESMES/CB for data reporting. Further investigations are needed to determine the model's scalability for a global context.

III.3.2 The SDMX initiative

38. A task force including the BIS, ECB, EUROSTAT, IMF, OECD, and the UN is working to identify business practices in the field of statistical information that would allow more efficient processes for exchange and sharing of data and metadata. This initiative has been called Statistical Data and Metadata Exchange (SDMX⁸). The goal is to explore common e-standards and ongoing standardisation activities that could allow gaining efficiency and avoiding duplication of effort in the field of collection and exchange of statistical information. The intention is to take advantage of existing and emerging exchange protocols, such as: GESMES; dissemination formats, such as that implicit in the IMF's Dissemination Standards Bulletin Board (DSBB); and e-standards, such as Extensible Mark-up Language (XML).

III.4 Using national source standard products, direct access to national source databases

39. The OECD relies upon NSOs standard products and, in particular on-line databases, as a way to reduce and to almost suppress the reporting burden on NSOs for data that are part of NSOs regular dissemination programme. The OECD has been using that approach for main economic indicators for many years. With the development of the web and of databases available on the web this method has been made even easier since no extra hardware and software is needed. The OECD uses this approach for the following countries:

- Web based database: Australia, Austria⁹, Canada, Denmark, Italy¹⁰, Mexico, Sweden and the United Kingdom.
- Non-web based database: Finland, France, New Zealand and Spain. For those we need specific software.

40. The advantages of using a NSO database for an International Organisation might include the following:

- Flexibility in terms of content, provided the database contains all the data needed;
- Control of timing of collection;
- Possibility of automatisation of data extraction;

- Highest quality data in NSO.

41. The approach of using databases is very efficient from the NSO's point of view. It could be made more efficient from the data receiver's point of view and applied to more countries if the following problems could be solved for those countries:

- Lack of completeness of coverage, in terms of subjects or time period;
- Different systems used in several parts of the same organization;
- Lack of support for metadata;
- Lack of reliability in systems used;
- Level of sophistication of some database systems making automatisisation difficult.

42. An example of on-line database from which data collection can be automated is Eurostat's New Cronos. The OECD collects data from New Cronos using a simple method by which Excel tables at the OECD are automatically refreshed with New Cronos data with only one click.

43. The disadvantage of using this approach is mainly one of scalability. The approach may be very practical for a handful of countries, but it would impractical to extend it to the full membership of the IMF, for example, as well as to multiple data providers in each country.

44. The web database approach might be made more scalable for international organisations if some standards for the on-line presentation of data and metadata could be defined and implemented by a majority of countries. This is one of the objectives of the SDMX group.

IV. Further increase the efficiency of the system, towards data sharing

IV.1 Global efficiency

45. On the receiving end of the exchange the main issues in terms of efficiency are those impacting on the quality of data and metadata and those impacting on the resources needed for the process. The relative importance of the various aspects of efficiency is dependent on the subject matter area. Timeliness, for example, is extremely important for short-term statistics but less important for annual surveys. The following is a list of attributes of data collection processes that have an impact on efficiency. Some of these attributes would be useful to elaborate in a generic form, such as a template, to facilitate the negotiation process, as a precursor to establishing data sharing arrangements. Additionally, these criteria may be applied to the overall process in deciding among different data exchange models, as described in section IV.2:

- Total processing time (from data receipt through processing);
- Reliability (error free, in statistical content or transmission);
- Flexibility (responsiveness to changes in data requirements);
- Repeatable (resource needs for high frequency requests);
- Cost of initial investment;
- Simplicity (IT and human resource requirements);
- Continuity.

46. Of course there is not a single type of process that is the most efficient for all criteria for all types of data collection by international organisations. Depending on the most important criteria a type of process will be the most appropriate. For example, in the area of short-term statistics, and where data have to be collected on multiple subjects, timeliness, flexibility and reproducibility are the most important criteria. Reliability is extremely important since errors in transmission, in data or in the process, imply multiple exchanges of data between provider and receiver(s) for the same data flow. From the OECD's experience this is the most important source of inefficiencies in the data collection process. It is inefficient for all receivers and also for the sender.

47. Given a number of attributes data exchanges can be efficient for individual areas but overall duplication or inconsistent use of technology, for example, might make the processes globally inefficient for the community of data providers and international organisations. Therefore, the efficiency of data collection has to be co-ordinated and seen globally.

48. The OECD has a decentralised structure for its statistics, as described in Section I of this paper. The co-ordination of data collection activities has been improved by the creation of a database for the OECD Statistical Work Program including details on data collection. This experience could be used as a model for the global co-ordination of data collection at the international level.

IV.2 Different possible models

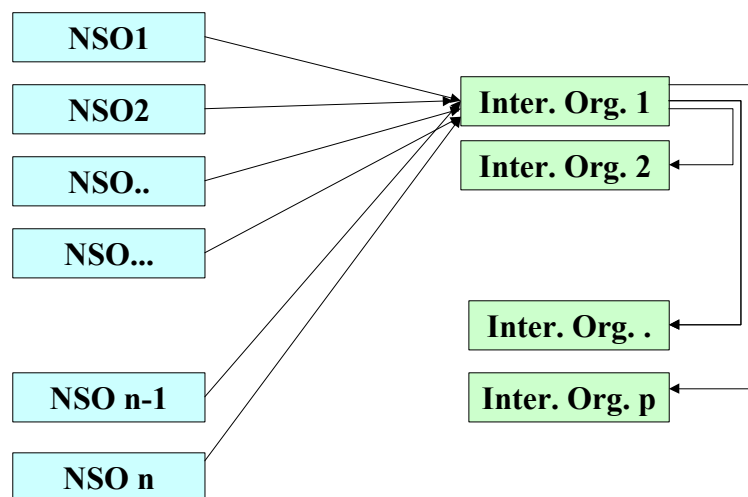
49. Data collection activities can be divided between those that collect data that are part of one country's dissemination programme and those that are special requests of international organisations and not part of a country's dissemination programme, for example the data collection for PPPs. In general the latter are well co-ordinated between the various International organisations involved. The rest of this section is devoted to requests that can be met by on-going data collection activities.

50. There are at least three possible models, as described below.

51. The first model corresponds to the situation where, for example, a NSO sends different files to all International Organisations involved in the exchange. In this case the reporting burden is maximum on the NSO. This first model represents the most common situation and is the only feasible model where data requirements, transmission formats and media are not common across international organizations. In the particular case where a NSO sends the same file at the same time to all International Organisations involved, this model represents the most efficient from the NSO's point of view. The model presupposes co-ordination of content, format and medium between International Organisations.

52. The next model represents the situation where data is sent to one International Organisation that redistributes the data to other organizations. It also presupposes co-ordination of content, format and medium between International Organisations.

Push model 2 - many to one and redistribution

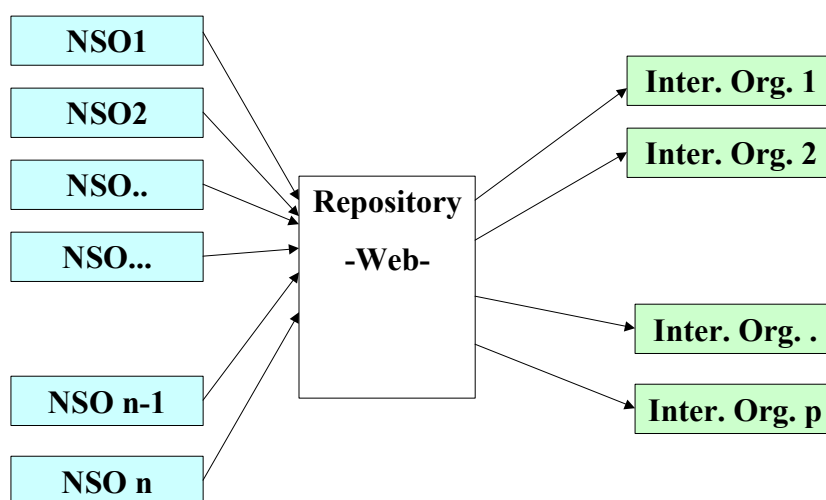


53. This model is currently used for the transmission of National Accounts data of EU countries and accession countries to Eurostat and the OECD. It also describes the situation with the ECB and IMF, where the ECB acts as a gateway to the IMF for the provision of financial statistics for all euro area countries.

54. A major drawback with this model is the lack of direct contact between International Organisations and national providers, other than the organization that is acting as the gateway. In this model it would be useful to have a mechanism to allow each International Organisation to verify that they have received all data files sent by the provider. Another problem occurs when a data error is discovered by one of the International Organisations.

55. The third model is the “data-sharing” model where, for example, NSOs upload their data to a repository that would be accessible to International Organisations. The repository could simply be the Internet. The data made available by each NSO need not use the same format or be in one physical database. If they would be in the same format, then it would be possible to present the data as a common database. A simple form of the data-sharing model is the situation where NSOs have all the required data available on the Internet in their web databases.

Data-Sharing model



56. This model is very efficient from the data provider's point of view and the receiving organisation would benefit from quality standards that data providers apply to their regular Web products. The other important advantages of this model are that it offers important opportunities for automatisisation and that it could be used as a common repository. Investigations should be made on a mechanism to publicize to all partners any corrections to data errors. Another area of investigation or experimentation would include how to use the repository as a virtual database with the development of a central catalogue. The main disadvantage to this model is that it might not be feasible in the short-term as it would require a high degree of cooperation and agreements among data providers and data collectors. Nevertheless, technologies such as XML would greatly facilitate the implementation of this model.

IV.3 A project for short-term statistics

57. A project is under way for co-ordinating the collection of data for major short-term economic indicators (MEIs) for the OECD, Eurostat and the IMF. They cover indicators such as Consumer Prices, Monetary Aggregates, Balance of Payments, Industrial Production, International Trade Totals, Labour Force, etc. All data needed by the OECD in this area are from NSOs' (or Central Banks') regular publication programmes.

58. From the OECD's experience using existing dissemination channels of data provider's is in general the most efficient way of collecting data. The current preferred approach for collecting MEIs at the OECD is to use on-line access to a database whether via the Internet or other means. Such access methods are in place for the NSOs and Central Banks of several OECD countries.

59. The IMF is exploring ways to make use of the data the OECD collects for *MEI* as a means of reducing the reporting burden on its member countries; in exchange, it hopes to

increase the timeliness of some of its short-term indicators for publication in *IFS*.

60. The following options were investigated for co-ordinating OECD collections of MEIs with those of Eurostat:

- Copies of files received by Eurostat from NSOs to be forwarded to OECD and the IMF;
- OECD and IMF extract from the Eurostat production database (EBT);
- OECD and IMF extract from Eurostat's New Cronos;
- Mixed approach where OECD and IMF collect from/with Eurostat when there is no database access at the NSO or Central Bank.

61. The last option is the one being currently pursued at the OECD.

62. The following issues are being addressed:

- Identification of MEI data required by the OECD in the flow of data sent by NSOs to Eurostat; a standard codification scheme would simplify this task;
- Identification of data required by the IMF in the set of data collected by the OECD; a standard codification scheme would also simplify this task;
- Metadata requirements;
- Ways of providing control and flexibility on requirements to the OECD and the IMF;
- Treatment of statistical adjustments: seasonal adjustment questions, trading days, history covered and historical revision policy;
- Consistency with other OECD and IMF data, respectively
- Resources and Systems problems for OECD, Eurostat, IMF and NSOs;
- Timeliness;
- Embargo rules and other legal considerations.

63. It is also proposed to use the study in the context of research on new technology by members of the SDMX group. The study is expected to deliver and implement a new business model for data collection for MEIs. The expected time frame for the project is the next 6-12 months.

V. Conclusions

64. Possible directions for further improving the co-ordination and efficiency of data collection by International Organisations, as described in this paper, are to:

- investigate the possibility of developing a common system for monitoring data collection activities by International Organisations;
- investigate the possibility of defining a standard codification scheme for economic time-series (per national accounts and balance of payments) with an initial focus on short-term indicators;
- investigate the possibility of defining a generic template to codify the basic elements of a

data sharing agreement, supplemented with “best practices”;

- encourage the development of data sharing models as defined in this paper;
- work with NSOs for facilitating the reuse of their standard products.

65. It is suggested to work with the SDMX to investigate and implement those ideas using new models and technologies in development in the market of data exchanges.

END NOTES

¹ Prepared by Ann McPhail, IMF Statistics Department, and Gérard Salou, OECD Statistics Directorate.

² Includes National Statistics Offices (NSO), Central Banks, Ministries and other national providers.

³ GESMES is an official UN/EDIFACT standard for the exchange of aggregated statistical data, see www.gesmes.org

⁴ See <http://dsbb.imf.org/>

⁵ In 2001 the OECD Statistical Program of Work included 98 statistical activities.

⁶ Extracted from the OECD Statistical Program of Work.

⁷ GESMES/CB has been renamed in GESMES/TS, with TS standing for Time-Series.

⁸ See www.sdmx.org

⁹ Austrian Institute of Economic Research.

¹⁰ Under development.

Glossary of Acronyms

BCEAO	Banque Centrale des Etats de l’Afrique de l’Ouest
BEAC	Banque Centrale des États de l’Afrique Centrale
BIS	Bank for International Settlements
BOPSTA	A GESMES profile for Balance of Payments Statistics
CSV	comma separated values
DSBB	Dissemination Standards Bulletin Board
EBT	Eurostat production database
ECB	European Central Bank
ECCB	Eastern Caribbean Central Bank (ECCB)
EDIFACT	Electronic Data Interchange
EIB	European Investment Bank
EU	European Union
EUROSTAT	Statistical office of the European Commission
GESMES	General Statistical Message (a UN/CEFACT standard EDIFACT message)
GESMES/CB	A GESMES profile for Central Banks
GDDS	General Data Dissemination System
IFS	<i>International Financial Statistics</i> , an IMF publication
ILO	International Labor Office
IMF	International Monetary Fund
MEI	<i>Main Economic Indicators</i> , an OECD publication
MEIs	main economic indicators
MS SQL	MicroSoft’s relational database software
NSO	National Statistical Office
OECD	Organization for Economic Co-operation and Development
PPP	Purchasing Power Parities
SAG	OECD Statistical Advisory Group
SDDS	Special Data Dissemination Standard
SDMX	Statistical Data and Metadata Exchange
SNA93	<i>System of National Accounts 1993</i>
UN/CEFACT	United Nations Centre for the Facilitation of Procedures and Practices for Administration, Commerce and Transport
UNACC	United Nations Administrative Committee on Coordination
UNECE	United Nations Economic Commission for Europe
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIDO	United Nations Industrial Development Organization
UNSD	United Nations Statistical Division
WAEMU	West African Economic and Monetary Union
WB	World Bank
WHO	World Health Organization
XML	eXtensible Mark-up Language