UNITED NATIONS		**E**

	Economic and Social Council	Distr. GENERAL CES/2002/7 5 April 2002 ORIGINAL: ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

<u>Fiftieth plenary session</u>
(Paris, 10-12 June 2002)

DELIVERY OF STATISTICAL DATA TO INTERNATIONAL ORGANISATIONS
A COUNTRY PERSPECTIVE

<u>Paper submitted by Statistics Norway</u>

INTRODUCTION

1.   The international system of official statistics consists of the National Statistical Institutes (NSIs) and several International Organisations (IOs). In general, the division of work is that the NSIs produce national statistics whereas the IOs collect and disseminate statistics from several countries and produce aggregates covering member countries. In order to achieve international comparability, concepts, standards and methods have to be harmonised. Thus, a main responsibility for international statistical organisations[1] is to be an arena for the work with common concepts, standards, and methods. The organisations can also give added value to national figures by systematic adjustment and harmonisation to ensure improved international comparability. This evaluation process is dependent on easy access to relevant metadata.

2.   Internationally comparable official statistics are increasingly important, both for national and international users. National statistics can in many cases only be understood and interpreted properly when compared to other countries, and more and more information is requested on supranational groupings both for political and business analysis. As a result of this demand, delivery of statistical information to international organisations – including Eurostat as a supra national body – has been gradually developing over the last 50 years, and is now at a high level, covering most statistical subject areas.

GE.02-

3.   Even if the focus of this paper is on some practical aspects related to data delivery to a variety of organisations, one should also consider the total effects of adapting to international requirements. On the positive side one can note that the use of international standards as some kind of "best practice" solution, may simplify the process of choosing methods and standards for use at national level. However, in some cases a strict adaptation of the different national statistics to international standards can make the national statistics less relevant for national usage. The international requirements can in those cases disturb the balancing process that is the obligation of an NSI, also taking account of the response burden and priorities in relation to available resources. Furthermore, as a result of the variable size of countries, an international standard (based on some kind of international average) for instance for accuracy, will result in an extra burdensome size of the statistical samples for small countries.

4.   For most NSIs answering requests from IOs represent a significant workload. Such requests may lead to extra frustration and difficulties, in particular in small NSIs, if:
- Standards and definitions are not well harmonised in relation to national principles;
- Standards and definitions are not well harmonised between several Ios;
- More or less identical requests come from different IOs (and there are transmission cost to each of them);
- There are arbitrary (not intended) variations in data claims from different Ios;
- IOs do not use electronic media for data transmission;
- Different IOs do not use the same technical solution for data transmission;
- The formats for data transfer are not well designed and efficient;
- Proper documentation/explanations are missing and previously sent information is not utilised.

5.   Thus it is important to focus on actions and initiatives that can further improve quality and efficiency in this transmission process. This paper will address some general issues, as well as some concrete cases, and hopefully, provide some ideas for further reflection and follow-up.

6.   It should be underlined, especially in the context of this conference, that many initiatives have been taken, resulting in improved coordination between the different organisations over the last 5 - 10 years. But the tasks of the organisations are increasing, new areas of international interest are developing, and there will thus be a continuous task to improve coordination of international reporting. The target can presumably be formulated as:

> The practical delivery of statistics to international organisations should place as little extra burden on national statistical institutes as possible. All requests should be well specified, well coordinated with requests from other organisations and using the most efficient transfer format. Internationally well-coordinated metadata will be a prerequisite for this.

SOME NATIONAL EXPERIENCES

7.   How the cooperation between the international statistical system and the national statistical system is managed will to some extent be dependent on how centralised the national statistical system is. If the national system is decentralised, requests from IOs can in principle go

via several channels, not only to and via the National Statistical Institute. In Norway, the Statistics Act gives Statistics Norway a role as coordinator of international statistical activity. And as the production of official statistics is rather centralised in Norway, we should be able to have a fairly good overview of statistics transmitted. However, there are some areas where the primary statistical production is carried out in ministries and directorates, and Statistics Norway at present only have a partial view of transmissions from these bodies.

8.   It should be noted that Statistics Norway has an active relationship to most international bodies, also Eurostat, as Norway is formally following most of the legal requirements of the European statistical legislation, based on the EEA agreement. It is estimated that more than half of the resources used for statistical production within Statistics Norway is directly or indirectly related to meeting requirements of the EEA cooperation. Thus the effects of international cooperation are much larger than the relatively minor resources that can be identified for the direct transfer of data (see below).

9.   Statistics Norway performed in 1999/2000 a survey of the delivery of statistics to different international organisations. The aim of this survey was to see how the internal organisation and processes for answering these requests could be improved.

10.   According to this survey (which does not cover all deliveries) it was indicated that there were some 80 regular deliveries to international organisations every year. Around 60 of these had a regularity of between two months and one year. It was further estimated that these deliveries required 3,800 man-hours of work, meaning around three persons on an annual basis. Compared to Statistics Norway's staff of 900, this overall burden does not appear high. However, as some requests may come in inconvenient periods (e.g. summer holidays) with short response time, there may be considerable frustration and problems to keep deadlines.

11.   The burden increases when the requests require manual extraction and adaptation, rather than automatic retrieval of a table or from a database. When assessing the extra workload, one should also take account of the fact that there are differences in the way people perceive and estimate this workload; some consider this as a part of the general processing of statistics, whereas others consider this as an extra work load.

12.   The international requests for statistics are considered to be difficult to handle when they are complex, covering many subject areas, with partly overlapping contents from different organisations – or even worse: based on somewhat different specifications and standards.

13.   The following is a general overview on which organisations data are delivered to –
related to subject area (following the structure of the UNECE Integrated presentation):

Subject area	Eurostat	OECD	IMF	UN	ILO	Nordic	Euro. Council	Unesco	WHO	WTO	ECMT
National accounts	X	X	X								
Money and banking, Gov. finance, BoP	X	X	X			X					
Foreign trade	X	X		X						X	
Transport and communication	X	X		X							X
Distributive trades	X	X									
Tourism	X								X		
Other marketed services (i.a. ICT)	X	X				X					
Price statistics	X	X									
Agriculture, fishery, forestry	X										
Industrial statistics	X	X		X							
Energy statistics	X										
Science and technology stat.	X										
Business register											
Demographic stat.	X		X	X		X	X				
Population and housing census	X			X			X				
Labour market	X	X			X	X					
Education and training	X	X						X			
Culture statistics											
Income and expend. welfare	X		X								
Social security											
Health (incl. causes of death)	X	X				X			X		
Gender statistics											
Social indicators and frameworks											
Environment statistics	X	X									

14.   In addition to the list of IOs, it should be mentioned that the European Central Bank
(ECB) to an increasing extent collects statistics on the financial market from National Central
Banks and NSIs, partly via Eurostat. The Bank for International Settlements (BIS) also collects
statistics, basically from National Central Banks. Furthermore, it should be noted that the
International Energy Agency (IEA) regularly collects energy statistics.

15.   In terms of resources directly used to compile the data to answer the request, it was
estimated that the deliveries within the following fields were rather demanding to the mentioned
organisations (some important deliveries might be missing):
  • Health statistics (OECD) (470 man-hours);
  • Agricultural statistics (Eurostat) (400);
  • Social statistics (Eurostat) (250);
  • Health statistics (NOMESCO (Nordic)) (250);
  • Education statistics (Joint questionnaire Eurostat, OECD, UNESCO) (200);
  • Environmental statistics (Eurostat and OECD) (200);
  • Demographic statistics (UN) (160);

- Demographic statistics (European Council) (140).

16.   It seems that the transmission process is functioning more efficiently for regular delivery of short-term statistics compared with annual or less frequent structural statistics. However, it should also be underlined that the resources used for international delivery are also dependent on how efficiently the data are organised at the national level. In Statistics Norway, the subject matter units until now have been responsible for international reporting within their field of responsibility. They have to use their regular production system and databases or publications to extract the data requested by the IOs. Thus, at the moment there is no consistent reference or dissemination database that can be used efficiently for data dissemination in all subject areas. This situation is not regarded as ideal and we have recently started with the development of a new structure of dissemination databases. Such databases will make the process of international data delivery more efficient.

17.   One important aspect related to efficiency of preparation and transmission is the media used. According to this survey in 1999/2000 the media used were the following:

	Total	Eurostat	OECD	Other
Paper (mail)	26	8	4	14
Telefax	2	1	0	1
Diskette	20	15	2	3
E-mail	39	19	13	7
Sum	87	43	19	25

18.   Even if transmitted in electronic form (that should be the target for most transmissions) the format used can have a major impact on how easy it is to prepare the extract (see the example related to the OECD). The format of the delivery will certainly also have an effect on how easy it is to use data on the other side – but this is not the topic of this paper. It should be noted that the area of data transmission is developing rapidly – with a rapid increase in electronic transmission – and the overview above is already partly outdated.

19.   An observation that can be relevant is that spreadsheets have often been used for data transmission. However, especially for structural, multidimensional data the design of these sheets will affect the way data have to be organised and the transmission process may be inefficient. Spreadsheets will normally have to be filled in by manual processes that are rather time-consuming. In other cases, they can be copied from other files – with several possibilities for error. For such transmissions it would be more efficient to have a well defined and agreed file format that can be produced from different types of local data bases.

TWO EXAMPLES: OECD SSIS AND IIS

20.   The deliveries of data to OECD Surveys on Structural Statistics for Industry and Services (SSIS) and on Industry and Services Short-term indicators (IIS) can serve as examples of some issues, both related to external and internal coordination, to standardisation of classifications and to formats of the data to be provided.

21.   The **SSIS** is requesting extensive structural statistics covering industry and services with a breakdown to 3 and 4-digit level of ISIC rev. 3. This is a joint questionnaire with the United Nations Industrial Development Organization (UNIDO) and is the prime source for the OECD Structural Analysis Industrial Database (STAN).

22.   It is our opinion that there are several problems with this questionnaire:
- The ISIC breakdown causes difficulties, as our production of statistics is based on NACE;
- The way the data are requested: one spreadsheet table with ISIC class in rows and with nine years as columns makes this awkward to handle;
- The delivery is rather similar to Structural Business Statistics delivery to Eurostat, and is accordingly felt as double work.

23.   We are aware that the coordination between Eurostat and the OECD on this specific topic is high on the agenda and discussed at the Statistical Working Party of the OECD Industry Committee.

24.   The **IIS** is asking for information on a quarterly basis and is now also extended to cover services.

25.   In our opinion this process suffers from many of the same deficiencies as the one on the SSIS:
- Use of the ISIC classifications makes the extraction of some figures not easy;
- One spreadsheet table with all ISIC classes and quarterly data for several years gives more work than necessary. (A simple text format would be easier, if each series is coded in a unique way. Lessons learned from the transmission of short-term statistics to Eurostat using GESMES/GESMES CB should be taken into account);
- There were problems with multiple correspondents and lack of internal coordination;
- This is felt to be double work as data are transmitted to Eurostat.

26.   Summing up this example illustrates that:
- Coordinated use of international classifications (in this case ISIC and NACE) is important;
- Organisation of the questionnaire format and the network of contacts are of major importance to improve efficiency;
- It is felt to be double work when delivering more or less similar data to Eurostat and the OECD – apart from different classifications.

SOME SPECIFIC ISSUES

27.   As stated in the introduction, international reporting is an important part of the activity in a NSI, and a lot of effort is put into the process of establishing international concepts, standards and methods and adapting national statistical products to suit the international claims. It is a pity, if the process of transmission and dissemination of international comparable statistics is a bottleneck, and is an obstacle for access to international comparable statistics. One part of this is the transmission from the national statistical system to international organisations. Another

equally important part is the access to data stored at the international organisation both from the NSIs and from other national and international users. Based on our experience the following are some issues that deserve attention in order to improve the efficiency of delivering data to international organisations.

Coordinated classifications/standards

28.   It is obvious that the NSIs will get much extra work, if international organisations do not use the same classifications, standards and code lists. Thus the organisations will have to agree on unresolved issues – or they will have to take more of the burden to harmonise data using a different classification. Based on the example: The OECD might have to accept data based on the NACE classification and will have to make adaptations to ISIC – if maintained.

Common and clear metadata descriptions

29.   Clear specifications of the data that should be provided, are fundamental in order to produce data in an efficient and correct way. The contents of such metadata specifications, should be agreed upon and follow some minimum standards. The descriptions should as much as possible be connected to, and integrated into the questionnaires, when there are ad hoc deliveries, whereas metadata descriptions for regular transmission only need to be agreed upon once and supplemented with revisions. This means that long Word documents with metadata and 'questionnaires' as spreadsheets are not considered to provide ideal solutions.

From passive to active transmission

30.   Many transmissions are based on the condition that data should be transmitted before a fixed date, as for instance set out in a legal act by Eurostat or based on gentleman's agreement. For regular transmissions it should in general be an agreement on a normal transmission date, which makes it easier to plan the deliveries. Any changes in the coverage should be pointed out before a certain deadline.

Structure of requests/transmissions

31.   Requests should be broken down to fairly logical entities – that make them easier to process in different organisations. Thus, huge requests for all short-term statistics or structural statistics are not operational, but should be broken down in manageable units. As these complex requests are often difficult to handle in centralised NSIs, one can expect that they are even more difficult to handle in decentralised systems. Such logical entities will also make it easier to provide common metadata descriptions and a common exchange format.

Transfer formats

32.   Organisations should agree on a common transfer format. As GESMES/GESMEC CB now is operational for Eurostat transmissions, and used for the major part of short-term statistics from Norway to Eurostat, this solution should be a strong candidate (see references and Annex for further description). Large, complex spreadsheets should anyway be avoided. The work going on in the framework of the Statistical Data and Metadata Exchange (SDMX) initiative initiated by the IMF and including the BIS, the ECB, Eurostat, the OECD and the UN, appears to provide a strong commitment in this direction (see also Annex).

Handling of new data and revisions

33.   Requests should only ask for updates or revisions of already transmitted data. Former data could be part of the feedback when asking for new data (cp the OECD example).

Coordination of common datasets

34.   Overlapping datasets that are required by different organisations should in principle be transmitted only to one organisation – if not agreed on a common transmission format that would make it possible for parallel transmission. For subject areas covered by Eurostat one could foresee that all members of the European Statistical System (European Economic Area plus gradually accession countries) only transmit data to Eurostat in these subject areas.

National and internal organisation

35.   In order to improve the overall efficiency of data exchange it is also important to review the national organisation of the statistical production system as well as the organisation within the NSIs. It can be of major strategic importance to develop reference and dissemination databases that are supported by standardised metadata solutions and standard tools for data exchange. Such databases might be centralised within a country or based on some common network solutions in a decentralised system. It is also important to put in place a network of contact persons responsible for different subject areas, and ensure proper coordination across the different subject areas.

Access to statistics stored at international organisations

36.   In order to improve the overall motivation for providing data to international organisations, and to enhance the quality of both international and national statistics it is essential that the NSIs and other data providers are given easy access to these data. First of all, the data providers should be given unrestricted access in order to check data, but it is also important that these data are provided in a format that allows usage and integration in national dissemination products, and that there are no formal restrictions for such integration. International comparisons provide important value-added to national statistics, and should be part of the normal exchange agreements. This condition will require a common pricing policy between the NSIs and the international organisations. If on one hand, the NSIs send data free of charge to the international organisations, it will be difficult to accept that international organisations will require payments for the return of the common data sets. In the case of Eurostat such exchange agreements with use of the data base New Cronos can be solved within the framework of the Eurostat Datashop agreement, whereas in the case of the OECD the access to statistical data is somewhat restricted or too expensive. The free access based on the OLIS database is considered to be cumbersome and inflexible, partly because it has a high level of security that makes it more difficult to use than the Internet. Also the cost of access to statistical data from the database Source OECD is considered to be too high.

CONCLUSIONS AND RECOMMENDATIONS

37.   A requirement for more efficient transmission at the national level appears to be the design and implementation of reference data bases, holding the data that are to be transferred to international organisations, including well structured and agreed metadata, and allowing for transmission of data in a standard form, supported by metadata.

38.   Some important points for follow up by the international organisations are:
  • To ensure that still unresolved issues related to the harmonisation of definitions and classifications are addressed and solved. This will be an important basis for common

solutions for data and metadata transmission;
- To continue the work agreeing on common channels and division of work. A country should only be requested to send equivalent data sets to one organisation;
- To put in place secure and operational methods and tools for the transmission of sensitive information. The experiences from the STADIUM/STATEL solution supporting GESMES CB should be considered in this context;
- To give both NSIs and other data providers easy access to statistics available at the international organisations in a form allowing comparisons and usage in national dissemination programmes.

NOTE

[1] By international organisations are meant all international organizations that have responsibility for a specific statistical programme or standard work, or are requesting statistical information.

References:

Andravisaneas, C. & Yancis, L.: Exchanging statistical metadata in the European system of central banks. Paper to Metadata Workshop 2001. Eurostat

Brüngger, Heinrich and Svein Longva (1998): International Principles Governing Official Statistics at the National Level: are they Relevant for the Statistical Work of International Organisations as well? Conference of European Statisticians 46th Plenary Session 1998, (CES/1998/22)

Coordination of statistical data collection from countries. Statistical Commission Thirty-third session, 5-8 March 2002. (E/CN.3/2002/30)

Ward, D. & Pellegrino, M.: Developing a common understanding of standard metadata components: A statistical glossary. Paper to Joint UNECE/Eurostat Work Session on Statistical Metadata, 6 - 8 March 2002.

ANNEX: FURTHER EXPLANATION OF SOME TOPICS

GESMES

1.   GESMES is an acronym meaning Generic Statistical Message. It has been developed by a group of European statistical organisations working within the international EDIFACT standards body. GESMES has all the features required to exchange multi-dimensional arrays and time series data, including metadata (and footnotes). It can also be used to transmit simple tabular data. The advantage of using GESMES, in preference to a proprietary data format, is that it is an internationally agreed standard, which is both open and fully functional. It is not tied to the format and constraints of one particular application. In particular GESMES supports the exchange of:
*   metadata;
*   multi-dimensional arrays;
*   time series;
*   administration data.

2.   In its generic form, it also allows the sending institution to use the data set structure it prefers. GESMES was accepted as UN/EDIFACT Status 1 messages in 1995 and was first published in the UN/D95A directory.

Requirements for using the simple GESMES format

3.   The type of array data that can be sent in the simple GESMES format can be any type of multidimensional data or table, but the following are pre-requisites:
*   the statistical concepts must be known to both sender and recipient;
*   the order of the dimensions and cells must be fixed and known in advance;
*   the coding system used for the statistical values must be known to both sender and recipient.

4.   More information on GESMES on: http://forum.europa.eu.int/Public/irc/dsis/eeg6

GESMES/CB

5.   GESMES/CB is a message designed for the exchange of statistical information between organisations in a platform independent manner. The message implements a **time series data exchange model** (GESMES/CB Data Model) which provides for the exchange of multi-dimensional time series and a variety of associated metadata. It employs an appropriate GESMES profile and EDIFACT syntax. Though GESMES is a generic statistical data model, which affords sufficient flexibility to describe syntactically virtually any statistical data model, GESMES/CB has a fixed syntax. This allows partner institutions to design and to build the applications needed to "read" and "write" GESMES/CB messages, avoiding intermediate files and special translators; the design of the read/write applications is further simplified by eliminating genericity which is not needed when exchanging time series data. Due to the fixed syntax, in most cases, the rules used in GESMES/CB are stronger and more restrictive than those

in generic GESMES. However, the current design allows the possibility of future enhancements and progressive generalisation, if this is needed, upon agreement of the parties involved.

6.   GESMES/CB offers several features:
*   Easy adaptation to any economic domain and flexible coverage of all types of economic - statistical data;
*   Modern representation techniques: conceptually 'clean' multidimensional keys;
*   Efficiency: it avoids the unnecessary repetition of information;
*   Series keys without length restrictions;
*   Easy introduction of new attributes;
*   Attributes at various levels (observation, time series and 'higher');
*   Exchange of 'rich' metadata, organised in a flexible and efficient manner;
*   Suitability for two-way exchanging time series: i.e. for reporting and disseminating;
*   A *paperless* dissemination of a whole statistical data base is possible: data, metadata, definitions, key structures and code lists can be electronically disseminated from the centre to other institutions;
*   Suitability for stage-by-stage implementation: it could start with the essential parts (carrying the administrative and the numeric data) and later the whole message could be used;
*   Easy implementation: no purchase of special software should be necessary;
*   Consistency with international standards (EDIFACT);
*   Wide user group: national central banks (NCBs) world-wide, Bank for International Settlements (BIS), International Monetary Fund (IMF), several national statistical institutes (NSIs), EUROSTAT-BoP, European Central Bank (ECB);
*   Consistency with the long-term goals of partner institutions (ESCB, BIS and EUROSTAT) for achieving interoperability.

7.   Further information on GESMES/CB can be found on www.ecb.int ("Statistics", "GESMES/CB").

GENERIC GESMES AND THE EDIFACT STANDARDS

8.   The goal of UN/EDIFACT-standardisation is to give to the exchanged data flow objects, a structure whose elements can be understood and processed by software applications without human intervention. The first messages were created for commercial data exchange environments and concerned rather simple and static object types like invoices, purchase orders etc. Since then new sectors, such as statistics, with more complicated and varied data structures have joined the industry and it became evident that certain message types would have to be made generic to satisfy user requirements of those sectors without actually exploding the number of messages.

9.   The statistical office of the European Union, EUROSTAT, who has lead the development of statistical UN/EDIFACT messages is implementing GESMES into the data flows between it and the Member States of the EEA (European Economic Area) and promoting the usage of the

messages by other international organisations and by other sectors.

10.   More information on EDIFACT on: www.unece.org/trade/untdid/

THE SDMX (STATISTICAL DATA AND METADATA EXCHANGE) INITIATIVE

11.   The BIS, ECB, EUROSTAT; IMF, OECD and the UN have joined together to focus on best practices in the field of statistical information that would allow more efficient processes for the exchange and sharing of data and metadata within the current scope of their collective activities.

12.   The goal is to explore common e-standards and ongoing standardization activities that could allow gaining efficiency and avoiding duplication of effort in own work and possibly for the work of others in the field of statistical information.

13.   The intention is to do this by taking advantage of existing and emerging:
  * exchange protocols, such as GESMES/CB which was implemented by central banks for exchanging time series;
  * dissemination formats, such as that implicit in the IMF's Dissemination Standards Bulletin Board;
  * e-standards, such as eXtensible Markup Language (XML).

14.   Further information on the project can be found on: www.sdmx.org. A description of the draft version of the SDMX' glossary of statistical terms can be found in Ward, D & Pellegrino, M. (2002).