

SEMINAIRE

E+; 3=! C

SEMINAR

STATISTICAL COMMISSION AND
 ECONOMIC COMMISSION FOR EUROPE



Distr.
 GENERAL

CONFERENCE OF EUROPEAN
 STATISTICIANS

CES/SEM.43/22
 18 May 2000

ENGLISH ONLY

Seminar on integrated statistical information
 systems and related matters (ISIS 2000)

(Riga, Latvia, 29-31 May 2000)

Topic III: Innovations in data collection and exchange

EDI AND STATISTICAL DATA EXCHANGE IN THE EUROPEAN SYSTEM OF CENTRAL BANKS

Submitted by the European Central Bank¹

I. INTRODUCTION

1. A key element in the statistical preparations for Monetary Union has been the introduction of efficient and effective means of statistical data exchange between partner institutions. Indeed, for almost two years the European Central Bank (ECB) and the EU national central banks, which together form the European System of Central Banks (ESCB), have already been using advanced and fully automated EDI facilities to serve their statistical data exchanges. The ESCB data exchange system is based on a special telecommunication network and a standardised statistical message format (GESMES/CB). The use of GESMES/CB is expanding rapidly, as more and more institutions in the statistical community realise the benefits of standardisation and integration in the exchange of statistics. The issues discussed here relate to the experience gained over these years regarding certain basic components (network facilities, contents, statistical modelling and metadata, message format) and their interaction. In addition, some conclusions are presented which could be useful for future statistical data exchange projects using EDI means and aiming at integration.

II. FACILITIES USED IN THE ESCB DATA EXCHANGES

2. Good data exchange systems contribute to efficiency, timeliness and quality. Easy automation and integration were primary concerns when designing the ESCB data exchange system. Simplicity was also considered to be very important, provided that functionality was not sacrificed.

3. GESMES/CB is the single message format used for the exchange of statistical data and metadata across the ESCB partners.

4. The ESCB statistical data exchanges are two-way and possibilities for integration are provided in both streams:

- statistical data reported by national central banks (NCBs) and some national statistical offices (NSIs) to the ECB; and
- statistical data regularly disseminated by the ECB to NCBs (and to the NSIs involved).

These data flows are based on file transfers (GESMES/CB messages) and the intention has been to allow partner institutions for these messages to build "loading" and "downloading" mechanisms to and from their in-house statistical database systems with the option of complete integration.

5. The network used provides acknowledgement facilities both for file delivery and (for the data reported to the ECB) application acknowledgement as well. The latter informs reporting NCBs that the ECB has "loaded" the reported data. This, in the event of errors in the statistical message sent, is replaced by an error reporting acknowledgement, which informs the reporting institution of the syntactical and semantic errors found in its GESMES/CB message.

6. In order to minimise the traffic over the network, statistical communication is based on an updates and revisions (U&R) principle. Reporting institutions are encouraged to report to the ECB only new observations (updates) and, for previously reported periods, any observations which have been modified (revisions). The same applies for the data dissemination originating from the ECB: the latter disseminates only new observations and, from the periods for which data have been disseminated in the past, only those observations which have been revised. The same principle applies to the exchange of qualitative information (attributes). However, from time to time, the ECB refreshes the databases of the partner institutions by disseminating CD-ROMs containing full databases, as they ensure the application of all updates and revisions which have taken place in a given period.

III. DATA MODEL AND METADATA

7. GESMES/CB has a powerful underlying data exchange model. This data model is time series-based and it was designed jointly by the Bank for International Settlements (BIS), the European Central Bank and a special Task Force. Its design was produced with a view to ensuring the fulfilment of as many statistical requirements as possible, but in a context in which automation could still be possible without huge implementation and maintenance costs for the partners involved. As in all GESMES implementations, it is also the case with GESMES/CB that the standardisation work on the message also encourages standardisation in some higher level (semantic) layers (e.g. code lists), a benefit which becomes even more evident for partner institutions.

8. In the GESMES/CB data model, metadata are split into structural metadata and attributes:

Structural metadata comprise the statistical concepts, code lists and the structure definitions (key families) which provide the basis for the regular data exchanges. Structural metadata are created, administrated, maintained and disseminated by the central institution managing data exchanges with partner institutions. In the case of the ESCB data exchanges, it is a unit of the ECB (Statistical Information Systems Division) which co-ordinates and performs these activities.

Attributes are those statistical concepts which are exchanged with a view to providing qualitative information. In the GESMES/CB data model, those attributes which will be used in the context of a data flow are defined within the framework of the underlying structural metadata (key family definition). As also explained earlier, partners do not need to report all possible attributes in all their messages. Once some of them have become known to the receiving institution (e.g. compilation method for a time series), these would not need to be re-reported in future, unless they were to change.

9. The maintenance of structural metadata is of extreme importance, since they form the basis of a data exchange and they also form the link to other data exchange circuits which may exist for similar data flows. For example, special care is taken with the structural metadata used in the exchange of balance of payment statistics, for which the world-wide use of GESMES/CB requires a high level of co-ordination among the central institutions involved (Eurostat, OECD, IMF, ECB).

IV. GESMES/CB: THE MESSAGE

10. GESMES/CB was, from the outset, a key element of the statistical work needed for Monetary Union and of the preparations for it. Moreover, since the start of Monetary Union, statistical requirements, and hence statistical data exchange requirements, have evolved rapidly; GESMES/CB has provided the means of meeting these demands efficiently and effectively. The message was designed jointly by the European Central Bank (ECB) and the Bank for International Settlements (BIS) in very close co-operation with the central banking community world-wide. Eurostat also offered very valuable GESMES advice and expertise with the aim of improving synergy and ensuring interoperability and consistency with its GESMES work vis-à-vis national statistical institutes. The ECB and the BIS have adopted GESMES/CB as their sole message format for exchanging statistical data with their central bank partners throughout the world and, of course, as their preferred message format for exchanges with other institutions.

11. Since May 1998 GESMES/CB has been used in a live statistical production environment to perform all two-way data exchanges (reporting and dissemination) within the ESCB. It has also been used with other international institutions and some national statistical institutes. The initial data exchanges (May 1998) were of balance of payments data, and since summer 1998 monetary data have also been exchanged. Since that time, the ECB has provided partner institutions with several additional data set structure definitions, and now hundreds of GESMES/CB messages are exchanged each month within the ESCB, covering several domains of economic statistics:

- Money and banking
- Interest rates
- Balance of payments
- Exchange rates
- Short-term statistics
- Consumer prices
- ESA 95 national accounts
- Government statistics
- Financial accounts
- Securities

These data exchanges are essential for the definition and conduct of monetary policy within the ESCB. The automation of the corresponding regular data transfers has resulted in a huge efficiency gain.

12. In several cases, if the data are already available in reporting institutions' databases, the conceptual and technical preparatory work needed to start exchanging

GESMES/CB messages does not exceed one day; and the technical description of the new data set (and metadata) structure need not exceed one page. This demonstrates that GESMES/CB provides the flexibility and efficiency essential for rapidly defining and describing data and metadata structures when new requirements arise.

13. GESMES/CB is appropriate for the exchange of any kind of time series (data and metadata); it also provides the means of describing the underlying structures in a completely paperless, platform-independent and syntactically standardised environment. The flexibility of GESMES/CB derives from the power of its underlying data model as earlier discussed; it uses the EDIFACT syntax but has been designed in such a way that the syntactical message structure is predefined and frozen. That is why GESMES/CB has proved very simple to implement. Most partner institutions have for example developed "read" and "write" applications that translate GESMES/CB to and from their statistical databases using in-house resources.

14. The central banking community in Europe (through the ESCB) and world-wide (through the BIS and its partners) would wish to see GESMES/CB become the sole channel for exchanging statistical time series data. This would make it possible to drastically reduce conversion costs and to enjoy gains in terms of interoperability.

15. Further information on GESMES/CB, and on the ESCB data exchanges in particular, is provided on the ECB's Web site pages for statistics (<http://www.ecb.int/stats/stats.htm>).

V. GESMES/CB AND THE CHALLENGES AHEAD

16. GESMES/CB is increasingly being used world-wide for the exchange of time series data and it is in this field that the basic challenge ahead can be found:

- testing and ensuring interoperability, when more than one or two data exchange circuits are involved and they need to exchange data.
- as before: testing and ensuring consistency, when different local data models for storage are used.

17. Reporting institutions usually express the concern that they report similar or almost similar data to different organisations using sometimes quite different means (e.g. different classifications and concepts, different formats, different reporting specifications, etc.). It is a field in which technical means to achieve interoperability can also facilitate and encourage to a great extent the harmonisation of some of the statistical layers.

VI. FACTORS FOR SUCCESSFUL STATISTICAL EDI IMPLEMENTATIONS

18. Using the experience and the lessons learnt over the last two-years with fully operational statistical EDI, the following points highlight areas which could be crucial to achieving successful implementations in statistical data exchange. In fact, the points below have been dominant considerations in the setting-up and maintaining of the GESMES/CB-based data exchanges in the ESCB.

19. The involvement of partners in decision making: this is very important for deciding what is technically feasible, which the optimal timetable and path over time of a stage-by-stage implementation are, which issues are problematic and for identifying those areas in which compromises can be made. Moreover, involvement in decision-making contributes in building commitments and close co-operation.

20. Central co-ordination in managing metadata within institutions: the benefits from using standardised statistical messages are even more tangible under a solid structural metadata framework with optimal conceptual solutions and maximisation of the use of commonly accepted concepts and code lists. It seems that central institutions should be encouraged to set up a central structural metadata repository and to staff it with people committed to looking for the best solutions in close contact with other institutions. This would make it possible to avoid duplication of effort on their part and on the part of reporting institutions. Also, in general, institutions should aim to accept as few data models for metadata as possible. Ideally, a single data model would allow better communication with users, lower costs for mapping and storing statistical concepts in different systems, easier integration and a smoother transition to an environment with standardised statistical data exchanges.

21. Responding automatically to all possible requirements for data exchange is not always feasible: trying to respond to all possible requirements which users may have can give rise to various difficulties and delay projects significantly. Also, this could lead partners to express reluctance if very detailed requirements which are not of essential concern add significant complexity to projects or if they require considerable changes to their systems.

22. Avoiding duplication of exchanged statistical information: one simple rule could be that whatever has been transmitted need not be retransmitted. Therefore, when integration and database-to-database communication are targeted, it is better not to use paper forms as the starting point for defining the data exchange guidelines.

23. Keeping the syntax of the statistical message simple: when feasible, it is better to allow complexity to be absorbed by the data model assumptions than to lead to a complex or insufficiently fixed syntax. The benefits from using statistical EDI messages are especially justified in an open community framework and when two-way data exchanges (reporting and dissemination) take place. Therefore, the higher the number of partners receiving statistical messages, the broader the benefits of a simple syntax.