

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

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (ii): Metadata interchange

The Franco-European EDIFLOW inventory: administrative metadata

Contributed Paper

Submitted by France ¹

Abstract

Since several years, INSEE has made up an inventory of data transmissions from France to international organizations. The content of this inventory makes up a start of horizontal use of administrative metadata. The paper describes what this inventory contains from "European" side and its French counterpart. It also points out the new expressed needs and a few elements of relationship with Gesmes and SDDS.

Depuis plusieurs années, l'INSEE constitue un inventaire des transmissions de données de la France vers des organisations internationales. Le contenu de cet inventaire constitue un début d'utilisation horizontale de métadonnées administratives. L'article décrit ce que cet inventaire contient du côté « européen » et sa contrepartie française. Il indique aussi les nouveaux besoins exprimés et quelques éléments de lien avec Gesmes et SDDS.

¹ Prepared by Bertrand MICHEL de PREVIA, National Institute of Statistics and Economic Studies (INSEE, France)

“The Franco-European Ediflow inventory : administrative metadata”

I. INTRODUCTION

A few years ago, INSEE had to, as Eurostat, identify the data which were exchanged between these two bodies.

The problem was posed by a quite simple slogan: 90/99. We wanted to send 90% of our data to Eurostat through electronic way by 1999. Everybody understands what the year 1999 means: on the other hand for the 90% the question has become : “ninety per cent ? 90 yes, but of what?” What were these sendings?

Their inventory and description were a matter of administrative metadata rather than statistical metadata. It is the experiment related here.

SITUATION

France produces statistical data in the framework of the System of Official Statistics which leans on the Institute itself, on Ministries Statistical System, and other producers (Bank of France, INSERM, CPSA, ...). It is a decentralized system.

INSEE operates directly as a producer or indirectly as a coordinator. Numerous data are brought together and “exported” towards international organizations.

The European sendings are defined by a legal basis or an informal agreement. Eurostat plays the role of data receiver and disseminator body.

Besides the inventory of exchanges is a matter of the EDIFLOW project, managed by the A2 unit of Eurostat in charge of the transmissions of European Statistical System (SSE). INSEE has worked since 1997 in partnership to establish it. It has allowed to get a provisional bilingual Franco-European version in 1999. This work gave then rise to a new data base which is today available on the Web.

This document leans on the experiment of the creation of the “Franco-European” product.

II. DESCRIPTION

OVERALL CONCEPTS

To describe these “data “exports” we have used some overall concepts. The first one is **Data flow**. A data flow is a (regular) sending of consistent data between a defined sender and an addressee. The flow exists because there is a request to which an answer is made.

The request leans on legal and regulation means: it is the **legal basis**.

One fulfils the sending by technical means : Eurostat recommendation is the use of **Stadium and Statel** software.

Either on Eurostat’s side (or of another international organization) or on producers’ side the actors of the process are persons (**intervening persons**) with roles , or bodies (administrative entities) in which and for which they operate.

CENTRAL AND EUROPEAN PART OF EDIFLOW DATA BASE

In most of the cases included in the directory, Eurostat is the receiver of our data. The starting point of a data flow is always the one of a request to which a producer answers, if possible.

What are the great categories of the “European” part of Ediflow data base?

The information related to the flows are the following:

- **Legal basis** of the request (which may be regulatory or an informal agreement)
- Brief **description** of the data flow and **statistical theme** which may be allocated to it, **working group** or steering groups associated.
- Target **statistical products and data base**
- Characteristics of the request : **format, confidentiality, periodicity and extension of time.**

The organisational and personal information will be symmetrically found with the requesters and the respondents.

The base contains the **name of the unit** of Eurostat which receives the request and the **name** of the person **in charge** of this unit with his(or her) **whereabouts** (telephone, fax, e-mail, postal address). These European central data are generally the same whatever the sending countries.

On the other hand **the person** - described by his (or her) **name - and his or her whereabouts** in charge of receiving the flow could be different according to the sending country, in as much the distribution is geographical.

An important part of these element can be visualised in Annex.

DESCRIPTION OF THE ANSWER

What are the characteristics of the data flow sent, that is of the answer?

The body which **produces the data**, acknowledged by Eurostat, is the one which transmits directly the data. It is called ANC (National Relevant Authority). The producer is identified in a general way by ANC and the basic unit by the field **CNA Unit**.

In the “European” base one must say which is the medium or **mean of transmission**, and the distinction is made between **physical medium** and **logical format**.

The national counterpart is more comprehensive because it contains non described elements - and it is normal - in the European base. This “national” information have been inserted, only for the flows emitted by our country. They are noted below.

The **mean** used for transmission and the transmitted **volume** were requested and must be described. In fact this latter information is too often absent.

Four fields describe in a concrete way the work performed.

Sending **follow-up** : it is noted if it exists (suivi)

Specific work : tells whether the international request involves a supplementary work

Work burden: quantifies the previous field response (specific work). Should lead to a burden in days/year. The idea was to assess the burden that the response represents for the producer.

The **particular constraints** (contraintes particulières) detail the content of supplementary work. It may correspond to a change of classification , requirements of extension of time,...

However there is no choice but to accept that the information was not sufficient to answer in a reliable manner when this question was posed officially and when a specific study was to be performed to assess the cost of SSE for France.

A field **evolution** allows to know whether the answer is stabilized or whether changes of the demand are programmed.

At last we have included two fields deemed subjective, but which are useful to us : **satisfaction of the requester** and **satisfaction of the producer**. The field of the satisfaction of the requester, at some moment contemplated has unfortunately disappeared from “European” forms. We carry on to put it in, because if the requester is not satisfied, evolutions must be expected.

The producer's is also significant even if remarks are not diplomatic. The subjective assessment are useful to coordinators and to producers but they can only be estimated with difficulty. When one points out that the statistics sent are probably useless because they are never published and that the producer is not pleased, it is meaningful. But are they administrative metadata?

At last we have added a **comment zone** in a working base newly created.

These elements can be visualized in annex.

III. DIFFICULTIES

DETAIL OF THE DESCRIPTION

The major problem we have to solve is the one of granulometry. What must be the detail of the description?

At the beginning, the Stadium transmission system and related Ediflow had to answer a simple question: was the expected file arrived? The unit was therefore the file transmitted by a country, it often was a diskette. The underlying idea was that one person might be able to manage these sendings.

The current evolution is much more pointillist. The answer can be broken down into tables. Thus we can have a flow, previously described previously as "main indicators of demography" turn into a list of 21 flows, each with a table showing variables and their mode of sorting.

Where to stop ? At the beginning 300 flows for a country appeared enormous!

In 1999 we had listed in a rather exhaustive way - after more than one year of work - 215 flows emitted by France toward Eurostat. In 2000 already Eurostat considered having passed from 245 listed flows to 521 (Report of LC 6 meeting). Today this figure has still increased

One can see here that the upper limit is the one of the elementary figure. But in that case, how to manage thousands (millions?) of transactions? The answer is perhaps the one of the regrouping by sets and to make the difference between a magnifying glass and a microscope.

However the administrative metadata would be almost identical - except for the volume - for one or a hundred files!

DIFFERENTIATION OF MULTIPLE NATIONAL RESPONSES

We have already described the form of response description with its fields, including in the French counterpart. Except for the language, the description of this response, for the European nucleus, even if it is less comprehensive, is correct. However beyond the form of the description, the issue of the content of the response can pose a problem.

The response of our country is a priori identical to the expected and preestablished European standard response. If the content is identical, the given description of the content is correct. On the other hand if the sent national content is different, then the standard European information is incomplete, and only the French national information is good.

The content of the response is not described in a detailed way country by country in the European data base. The response is assumed to be true to the request. This comes from the origin of the data base more centred on the transmissions independently from the content of the sendings. Paradoxically, the Ediflow system does not integrate the divergences or exemptions in relation to the norm, since it does not describe, in these cases the detail of the file sent by each of the countries. If the flow is broken down into two, there will be two sub-flows and the differentiated detail of their content only appears on the national side (at least in the Franco-European version).

LINKS WITH THE PRODUCERS

The Franco-European Ediflow inventory documents the statistical exports of France toward Eurostat. It can be extended to exports toward other bodies. But it is only the inventory of transfers.

This classification of exports does not cover the production and the link with the sources.
In this sense it is not thorough and it is not the single information system of French Statistics.

IV. GESMES & SDDS

CORRESPONDENCE

The first generation of Ediflow could describe in a simplified way the exchanged data flows. But the evolution of the descriptions toward more detail has often shifted from a general statistical domain, to a table of elementary variables. If the first level is sufficient to know the persons in charge the second only allows to know whether all the data were received or not.

At what time does the junction with Gesmes arise ?

One could say in a simplistic way that the data flows correspond to the notion of groups of "key family". But which is the correspondence with each of the key-families ?

It is at the level of elementary data that one can have a correspondence between the detailed flows, described recently and the series transmitted with Gesmes.

ROLES

It can be interesting to compare the 4 roles defined in Gesmes with those of Ediflow.

Gesmes :

- 1)CC - responsible person for information production
- 2)CP - responsible person for computer data processing
- 3)CF - Head of unit for information production
- 4)CE - Head of unit for computer data processing

Ediflow :

a) Receiver Roles:

(Data) Receiver : Role of person (within Eurostat) responsible for receiving statistical data.

Requester : Empowered person within Eurostat; this is the person to address whenever clarification is required.

b) Sender Roles:

(Data) Sender : Role of person responsible for sending statistical data to Eurostat.

(Data) Producer : Role of person responsible for the production of a dataflow.

Statistical Working Group Delegate : Role of person responsible for the request at the CNA; this is the person to address whenever there are problems and clarifications are required.

Production Co-ordinator : Person co-ordinating the production of a set of dataflows.

Daily Contact Person : The person which is the day to day contact point for a statistical data manager at Eurostat; this information will be maintained by Eurostat.

c) For dataflow roles and administrative positions: Collaborator

One can say that the roles defined around Ediflow are richer, nearer to the administrative structure and to reality of today. But only the precision of the definition of tables and validated codes will allow perhaps to integrate them one day in Gesmes.

DESCRIPTIONS OF SDDS

In SDDS the country field is the first one ; the geographical coverage is self-evident, but it may also correspond to a set such as Europe of 15 Members

The date of the latest update of the DSSB (Data Dissemination Bulletin Board) corresponds to the date of value of the metadata : this piece of information is not explicitly indicated in Ediflow.

Contact person : corresponds to the notion of « person to contact » of the Ediflow database, but with no distinction of role; this piece of information, a very useful one, implies however in SDDS an idea about the consultants, important and little frequent; because they are the hierarchical persons in charge of production which are often mentioned, now they should not be bothered at any moment.

Dissemination medium: in SDDS the description is more thorough and realistic than the one of Ediflow focused on transmission.

Dissemination lag of time : very useful and explicitly mentioned in the regulations of European Union, but not always indicated in Ediflow in an intelligible way.

These descriptions are fit to the needs of dissemination on paper or through the Internet. However, they stand out as a de facto standard. This one may be completed by a refined use of Ediflow concepts.

V. CONCLUSIONS

This experiment of partnership between France and Eurostat has not been re-used - for the time being - by other countries.

The updating difficulties did not lack.

This quick review of the application fields of the "European" application Ediflow and its French match, show that they cover well the description of structures and persons in charge of production and transmission, of forms of transmissions, but that the description of the data themselves is more complex.

The Gesmes fields are the minimum required for an optimum transmission, above all in Edifact. They do not cover the same needs as those of Ediflow data base.

A comparison between these descriptions and those of SDDS remains to be done notably in the framework of SDMX (Statistical Data and Metadata eXchange).

The overall integration of information on the data, their environment and their transmission is a large undertaking which remains to be taken on.

At last, we have in France a coordinating committee (Quality Label committee) in order to avoid to repeat requests already made with individuals and enterprises to prevent them from an increased burden. Ediflow tool, even as a draft, is relevant to analyze the doubles of requests of international organizations to National Institutes of Statistics and other producers.

EXAMPLE OF FRANCO-EUROPEAN EDIFLOW DATABASE; 1999 EDITION

EDIFLOW DIRECTORY SERVICES - [Dataflow inventory]

Theme: 4 Eurostat: Energy and Industry Request: SBS-IND Structural Business Statistics of Industry

Unit: D2 - Eurostat: Structural business statistics Request information

Head of Unit: BERNARD LANGEVIN Requester: OLIVIER MICOL

Dataflows identification

Dataflow	Dataflow description	Periodicity	Deadline
SBS-IND-A	Structural Business Statistics of of Industry	Annual	T+18 Months

Transmission characteristics of: SBS-IND-A

State	Competent National Administration	Mean trans.	Volume	Role of the people involved
Finland	STFI - Statistics Finland	Diskette		Benoît DE LAPASSE (Producer)
France	INSEE - Direction Générale: Institut National de la Statistique et des Études	STADIUM		M. Bertrand MICHEL DE PRÉVIA (Local coordinator)
France / 2	INSEE - Direction Générale: Institut National de la Statistique et des Études	STADIUM		
Germany	SB - Statistisches Bundesamt	Diskette		

Print preview Print to printer

People's involved roles

Démarrer Libération.fr - M... Boîte de récep... art_mtdws_030... art_mtdws_030... EDIFLOW D... 13:07

EDIFLOW DIRECTORY SERVICES (Dataflow information)

Extended Transmission Characteristics Information

State	FR	France			
Request	SBS-IND	Structural Business Statistics of Industry			
Dataflow	SBS-IND-A / 2	Structural Business Statistics of Industry			
CNA	INSEE	Direction Générale: Institut National de la Statistique et des Études Économiques.			
CNA Unit	INSEE-DGE2-E220	Div. Synthèse des statistiques d'entreprises			
Code/Description Internal	EU98D-E210A	4 fichiers correspondant aux séries 2N, 4N, 3A, 1D			
People involved	Benoît DE LAPASSE (Producer)	Requester satisfaction: correcte Producer satisfaction: nd			
Evolution	Modifications importantes récentes liées à la mise en place du règlement sur les statistiques structurelles relatives aux enl				
Particular constraints	Délais à respecter: toujours vrai				
Work effort	5 jours				
Specific work	Oui - unité	Follow-up: Contact téléphonique			
** Mean transmission	STADIUM	** Volume	Response date	08/04/99	
* Mean transmission		* Volume	File name	IND1AA	
** Data format	Other	* Physique Support	* Logical format		
Stadium station	FR-INSEE-02	Stadium application	FR-INSEE-SBS	Consignment	SBS-IND1-A
Status		Stadium mode			
Comments					

** EUROSTAT * CNA

Mode Formulaire FILT NUM

Démarrer Libération.fr - M... Boîte de récep... art_mtdws_030... art_mtdws_030... EDIFLOW D... 13:08

DEMANDE

Demande DEM-EU02P-SBS-IND

Dernière modification : 20/2/2003

Titre : **Structural Business Statistics of Industry**

Libellé demande(producteur) : **Stats. Structurelles sur l'Industrie**

Libellé demande(demandeur) : **Structural Business Statistics of Industry**

Code de la demande(demandeur) : **SBS-IND**

Confidentialité :

Thème statistique : **Eurostat: Energie et industrie**

Groupe de travail concerné : **Structural Statistics of Industry and Small and Medium-sized Enterprises**

Produits statistiques : **Mise à jour de la banque de données CRONOS (INDE); Publication "Structure et activité dans l'industrie"; "Panorama of European industry" annuel; "Structure and activity of industry" annuel ;**

Base de données concernée : **Databases INDE and DAISIE**

Format : **Seq. file**

Support demandé :

Périodicité : **A**

Délai :

Evolution : **Nouveau règlement procédure stabilisée**

Satisfaction du demandeur :

Commentaire : Unité dde : EUROSTAT-D2

Réponses :SBS-IND-A

"Cette demande est répondue pour les indices par l'INSEE et pour les données par le SCEES(à contrôler)

Base légale : BAS-SBS-IND

Titre : **Stats. Structurelles sur l'Industrie**

Libellé détaillé : **Directive du 6 juin 1972 (72/221/CEE) sur les statistiques annuelles sur l'activité industrielle + Règlement du Conseil 97/58 du 20/12/96 relatif aux Statistiques structurelles sur les entreprises (Séries 2A, 2B, 2C, 2D, 2E, 2F)**

Organisme demandeur : ORG-EU-EUROSTAT-D

Dernière modification : 8/1/2003

Nom de l'organisme : **EUROSTAT-Direction D**

Identifiant organisme : **EUROSTAT-D**

Libellé Niveau 1 : **6 EUROSTAT**

Libellé Niveau 2 : **Statistiques des entreprises**

Libellé Niveau 3 :

Adresse : **Bech B4/303 Bâtiment J. Bech - 5, rue A. Weicker**

Code postal : **L 2721** Ville : **Luxembourg-Kirchberg**

Téléphone organisme : **00 352 4301 35474**

FAX organisme : **00 352 4301 35409**

E-mail organisme :

Adresse internet organisme :

Commentaire

Unité de réception : ORG-EU-EUROSTAT-D2

Dernière modification : 3/2/2003

Nom de l'organisme : **EUROSTAT-Unité D2**

Identifiant organisme : **EUROSTAT-D2**

Libellé Niveau 1 : **EUROSTAT**

Libellé Niveau 2 : **Statistiques d'entreprises**

Libellé Niveau 3 : **Statistiques structurelles des entreprises**

Sigle :

Adresse : **Bech B2/303 Bâtiment J. Bech - 5, rue A. Weicker**

Code postal : **L 2721** Ville : **Luxembourg-Kirchberg**

Téléphone organisme : **00 352 4301 37286**

FAX organisme : **00 352 4301 32600**

E-mail organisme :

Adresse internet organisme :

Commentaire

Intervenants :

Rôle	Nom-Prénom	Téléphone	Fax	E-mail
Receveur	Mr. LIENHARDT Jean	352 4301 33856	352 4301 32600	Jean.lienhardt@cec.eu.int

REPONSE

Réponse REP-EU02P-SBS-IND-A_INSEE-DGE2-E220A

Dernière modification : 9/12/2002

Titre : **Activité et investissements des entreprises industrielles**

Libellé réponse producteur : **Activité et investissements des entreprises industrielles d'au moins 20 salariés; annuel(résultats définitifs)**

Code réponse producteur : **EU99P-E220A**

Code réponse demandeur : **SBS-IND-A**

Libellé réponse demandeur : **Structural Business Statistics of of Industry**

Format : **Seq. file**

Support :

Volume en Koctets :

Périodicité : **A**

Mode de transmission : **S**

Suivi : **1**

Travail spécifique :

Charge de travail spécifique : **30**

Contraintes particulières : **1**

Evolution :

Satisfaction du producteur : **nd**

Commentaire

CNA Unit :INSEE-DGE2-E220

Request :SBS-IND

Unité répondant : ORG-FR-INSEE-DG-E220

Dernière modification : 22/1/2003

Nom de l'organisme : **INSEE-DG-Div. Synthèse des statistiques d'entreprises**

Identifiant organisme : **INSEE-DGE2-E220**

Libellé ANC :

Libellé Niveau 1 : **Direction des Statistiques d'Entreprises**

Libellé Niveau 2 : **Département Système Statistique d'Entreprises**

Libellé Niveau 3 : **Division synthèse des statistiques d'entreprises**

Sigle :

Timbre : **E220**

Adresse : **18, Bd Adolphe Pinard**

Code postal : **75675**

Ville : **Paris cedex 14**

Téléphone organisme : **33-1 41 17 50 70**

FAX organisme : **33-1 41 17 65 20**

E-mail organisme : **dg75-e220@insee.fr**

Adresse internet organisme :

Commentaire

Intervenants :

Rôle	Nom-Prénom	Téléphone	Fax	E-mail
------	------------	-----------	-----	--------

Transmission par stadium : STD-SWC1-SBS-IND-A

Dernière modification : 17/2/2003

Envoi : **SBS-IND-A**

Station : **FR-INSEE-SWC1**

Groupe de flux : **FR-INSEE-SBS**

Identification interne fichier : **SBS-IND-AA**

Contrat d'envoi : **1**

Références contrat (product) : **en cours**

Date : **04/12/2002**

Références contrat (transmett) : **en cours de signatur**

Date :

Commentaire

28/1/2003, BMP au CNIO : Envoi 2/10/2002 : la machine propose 2003-1 ce qui est correct;

Le contrat a été rédigé pour le flux définitif SBS-IND envoyé par Philippe BARBIN (E220)