

NATIONS UNIES

COMMISSION ECONOMIQUE  
POUR L'EUROPE

ОБЪЕДИНЕННЫЕ НАЦИИ

ЭКОНОМИЧЕСКАЯ КОМИССИЯ  
ДЛЯ ЕВРОПЫ

UNITED NATIONS

ECONOMIC COMMISSION  
FOR EUROPE

SEMINAIRE

СЕМИНАР

SEMINAR

STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN  
STATISTICIANS



CES/SEM.35/SV/R.1  
19 July 1996

ENGLISH  
Original: ENGLISH and FRENCH \*/

Seminar on Official Statistics - Past and Future  
(Lisbon, Portugal, 25-27 September 1996)

**SESSION 5: OUR LEGACY TO FUTURE GENERATIONS**

**ELECTRONIC DOCUMENTATION OF THE STATISTICAL OPERATIONS:  
A FRENCH-ROMANIAN EXPERIENCE**

Report submitted by the National Institute for Statistics  
and Economic Studies (INSEE, France) and the  
National Statistical Commission (Romania) \*\*/

1. The concern of the National Institute of Statistics and Economic Studies ( **INSEE** ) for the documentation of its statistical operations arose fifteen years ago as a double concern of transmission of know-how within the Institute and of search for consistency in data processing. Mr. Edmond MALINVAUD who was General Director of **INSEE** decided to initialize the works on this topic by entrusting Mr Claude POULAIN with that responsibility. His research lead him to the developing of an instrument which was named **DDS (Statistical Data Dictionnary)**.

2. This instrument has to be looked at from two sides: first as an organization method of the documentation of statistical operations, secondly

\*/ As the document was submitted after the 10-week deadline, its distribution prior to the Seminar could not be guaranteed by UNOG services in Geneva.

\*\*/ Prepared by Dominique Crosnier, INSEE, France and Liana Mariana, National Statistics Commission, Romania.

GE.96-

as a computerized tool which assures its concrete implementation. At first, this instrument has been dedicated to statistical surveys, a type of operation essential to **INSEE**. In this context the point of view of producers was favoured, because it was a question of formalizing first of all the documentation to be used in the management of this type of operation. A data

model has been designed for it (Survey model) and a computerized tool has been associated with it (SURVEY). Nevertheless the disseminators came quickly with their own needs which has led to the definition of another data model aimed at representing a simplified documentation of the statistical operations from their sources to their methods of dissemination, another computerized tool being associated to that model: ISIS (Inventory of the Statistical Information System). Finally last a first attempt of generalization was started (T EXO). It allows, within some limits, to adapt the SURVEY data model to other situations. Practically, it has been used for the designing of documentation retrieval bases.

3. More recently, since 1992, in the framework of a cooperation programme between **INSEE** and the National Commission for Statistics of Romania (**CNS**), the latter has expressed its interest both for this idea of documentation of statistical operations and for the instrument developed by **INSEE**. A cooperation scheme between both Institutes has therefore been undertaken on the subject. It has been financed by European Union through the **PHARE** programme.

4. First, **CNS** could become initiated into the tools set up by **INSEE**, through the installation in Bucarest of the Survey tool in a micro-computer version. Then it has made some developments from existing tools (1994). Today, it has decided to carry on the collaboration with **INSEE** aiming at the implementation of the evolutions of this instrument.

5. This document intends to make a balance of this whole operation by dealing with the three following points:

1. A presentation of the **DDS** instrument, limited to the "Survey model" and the Survey tool which is associated with it. It is not, useful to consider the different models and existing tools to describe the principles of **DDS** and, moreover, only the Survey tool is available both at **INSEE** and **CNS**.
2. The achievements made **INSEE** and **CNS** with that tool.
3. The developments to be forecasted in the light of the concrete experiences.

#### **I. THE DDS UNDER THE FORM OF THE "SURVEY MODEL"**

6. Under the form of the "Survey model", **DDS** is first an organization method for the documentation of statistical surveys. It also has an operating version in the form of a computerized tool.

##### DDS as a documentation method

7. It is a method because it relies on a conceptual model for metadata (See **Annex 1**), with the help of which one organizes the documentation of a statistical survey in the following manner:

- Definition of the "objects" called " **types of form**" which allow for the description and classification of the different items which are useful for the implementation of a statistical operation in its phases of design, realization, operation, dissemination.
- Description of the links between these different items.

The purposes of this method are multiple :

- To build up and use the documentation during the process of compiling of the statistical survey. It is particularly necessary to make the dissemination of information between the different actors homogeneous.
- To store the memory of a survey in a "unique computerized place".
- Use this memory to inform, to design a new operation, to prepare archives or to disseminate the metadata.

#### The types of form of the survey model

8. The "Survey model" defines ten main "types of form"

- |                 |                  |
|-----------------|------------------|
| . variable      | . incident       |
| . questionnaire | . program        |
| . question      | . classification |
| . concept       | . table          |
| . specification | . file           |

9. In the center of the model one finds the type of form "Statistical variable". The remaining types have a close relationship with this, be they upstream or downstream.

#### Upstream :

- The **questionnaire**, reference document and the **questions** from where the variables most often originate.
- The **concepts** which are basic to a variable and clarify its definition.
- The **specification** notes which make precise the collection procedures and /or the algorithms of information processing.

#### Downstream :

- The **incidents** where one will be able to store all the comments linked to the whole processing, should they happen during the collection or not.

- The sources of the computer **programs** for checking, correcting or generating the computed variables.
- The composition of the **tables** used in the dissemination of the results.
- The description of the computer **files** built during the operation.
- The **classifications** used to classify the units observed by the statistical operation.

#### The fields of the types of form

10. Each type of form is described by a set of properties. Some are common to all types of form, others are specific to some of them. This set makes up the type of form. **Annex 2** gives the list of these properties called fields.

#### The links

11. A priori, reciprocal and links of several types may exist between all the types of form. In the implementation form of the model, it has been chosen to simplify this theoretical state, retaining only a limited number of link possibilities (See **Annex 1**) and organizing them into a hierarchy in such a way that a link can always be established between any two types of form even if it is indirect and passes through a third type of form. Therefore, there will exist "upstream" links and "downstream" links, "direct" links and "indirect" » links between the different types of form.

#### The storing mode of the documentation

12. From this information structure, any text document will be possibly registered in a **form** which will be the unit element of the database. A form is made of three parts. The first part describes, by its property values, the form to which the documentary text corresponds. The second part of a form is made of the documentary texts themselves: contents of a note, wording of a question, definition of a concept, source or reference of a computer program, table of the modalities of a variable, report of an incident, source of a table query, set of the items of a classification, descriptive text of a file. Lastly in the third part, one describes the links between a specified form and other forms of the database. **Annex 3** gives an example of a form of variable type.

#### DDS as a computerized tool

#### Technical aspects

13. The Suvey tool exists in three versions :
- A version under MVS in which the word processor is the one of ISPF, the DBMS being ORACLE. It has been developed with version 5 of ORACLE. Maintenance has been done in the versions 6 and 7. It uses

the DCF document designing software and contains an interface with SAS. This version has been installed in **INSEE**.

- A single station micro version in which the word processor is WORD 5.5, the DBMS ORACLE. The implementation has been done with version 5 of ORACLE. This version does not include the SAS interface, and has been installed at **INSEE** and **CNS**.
- A local network version for micro under UNIX in which text processing is WORD 2.0. and DBMS is ORACLE. This version has been developed with version 6 of ORACLE. It does not contain either the SAS interface. This development has been done in collaboration between **CNS** and **INSEE**.

#### The organization

14. There are three types of actors:

- **The Oracle database administrator** is the only person authorized to open a database, to register the managing clerks and to make the backups of the whole database (contents and ORACLE elements).
- **The managing clerks** are declared, by a symbolic name, to the database administrator and they can do all the management operations on the database, including backup in the meaning of the **DDS**, that is to say only the contents of the database.
- **The persons in charge of the query** are not specifically declared to the database administrator. They have the access to query the database or to recover information from the database. In no case they can modify the contents. Any person having a computerized identifier can potentially query the database.

#### The main functionalities

15. Depending on the case, they are available to all or only to the managing clerks:

Operation	Actor
Management	Managing clerks
Query of the database	Managing clerks and Query persons
Editing, recovering of data	Managing clerks and Query persons
Codebook	Managing clerks
Interface with SAS	Managing clerks

Management

Form loading can be done in batch or in "on line mode", either for the introduction of forms or for their modification. A consistency check allows verification of whether the introduction has been made according to the tool management rules. Introduction of texts (or their modification) can be done

directly in the ORACLE database or using the word processor ( ISPF, WORD 5.5 or WORD 2.0 depending on the case). This option is particularly useful when a text is recovered as a computer file.

16. The other main functions for management are the following :

- Deletion of forms.
- Addition of links, modification of the links order.
- Change of the name of a form.
- Introduction and/or correction of keywords.
- Archive (forms, historical review).

17. This last function may be particularly important. One can decide, at the time of elaboration of the statistical surveys documentation that it is no longer useful to save a textual document in its comprehensive link to other parts of the documentation (fields values, links), but that it is interesting however to keep a trace of it. For instance there might be an intermediate version of a specification note. One has in that case the possibility to keep it in archive where it will be stored without its links neither its fields values other than its title. The historical archive, may be realized on a computerized medium by unloading the contents of a database sorted by type of form.

Query of the database

18. There are several ways of querying the database:

- Form name selecting (when the names are known).
- Form type selecting.
- Selecting on the different fields of the types of forms. Depending on the case one will have closed choices (list of a criterion modalities) or open choices (selecting on a string of characters) for example on the name of an author or on value brackets (dates).
- Selecting on strings of characters contained in the text.
- Selecting from links between forms.
- Selecting on keywords.

Editing and recovery of data

19. There are two main standard states: editing of forms or editing of text. In the case of form editing, one gets the contents (the value of the different criteria) corresponding to the descriptive part of the form, together with the text. This edition is only available on paper. In the case of editing the only text linked to a form, one can edit into a file or on paper. Forms can be recovered, partly or wholly, either to prepare updates or to transfer them into another **DDS** database.

Codebook

20. A standard output allows editing into a file, then on paper, the list of the survey variables with:

- The symbolic names of the variables.
- The text part of the variable form with a full title.
- The list of modalities associated, and of comments which can include any type of references, even conversion tables.

21. If moreover the survey is exploited with SAS software, one can, in the MVS version, add to it the counts and the corresponding frequencies for each of the variables. The variables can also be grouped by themes.

#### The interface with SAS

22. Besides the previous possibility to introduce automatically distributions into the codebook, and always in the case of the MVS version, the interface with SAS allows to make some checks, two of which are particularly useful. An "exhaustive check" allows to verify that all the variables of a SAS table are also described in the **DDS** and that all the variables declared in the **DDS** really exist in a SAS table.

23. A "type check" which allows to verify that the values taken by a variable really belong to a list of allowed values. In particular, editing will be possible, with a classical PROC FREQ, of the values found for a variable in a SAS table, whereas they are not allowed for in the **DDS**.

## II. THE ACHIEVEMENTS

### The French achievements

24. The SURVEY tool being as a priority dedicated to statistical surveys and more particularly to the households surveys, this area has experienced the most achievements. As, to this day, there is in **INSEE** no obligation to make a documentation data base and as consequently this work is done on a voluntary basis, all the operations of **INSEE** are not provided with it. Nevertheless, most of the "regular" household surveys have made at a moment or another a **DDS** base, the most notable exception being the Population Census. On the other hand the use of this type of method remains embryonic in the area of business statistics and for regional surveys. Besides, a few still timid attempts are being achieved for operations the statistical source of which is different from "classical" surveys. It is the case, for instance, of the consumer price index.

25. It is advisable however to note that when an operation has been conducted as a partnership, and that **INSEE** had committed in doing the documentation of it, then the units in charge of these operations at **INSEE** have used **DDS** bases, in order to put to disposal of the different partners an homogeneous and consistent documentation (Transport, Health, Social Mobility and Integration, Leisure surveys).

26. Apart from the achievements made with the SURVEY tool, two other bases deserve to be quoted: the documentary base of the National accounts, made with TEXO tool and the ISIS base.

27. To be more precise, we shall quote the main existing bases and the main part of their contents.

The main existing bases

28. In the area of household surveys:

The set of the **Employment** surveys for 1980 and 1990.

- The set of the years 80 contains 739 forms among which 569 variables and 28 files.
- The set of the years 90 contains 660 forms among which 498 variables, 9 concepts and 13 classifications.

The **Training and Professionnal Qualification** surveys

- The base of the 1985 survey contains 1783 forms among which 549 variables and 248 programs.
- The base of the 1992 survey contains 2024 forms among which 1180 variables, 124 programs and 447 tables.

The **Housing survey**

- The base of the 1988 survey contains 929 forms among which 559 variables, 119 tables and 6 classifications.
- The base of the 1992 survey contains 1024 forms among which 828 variables and 33 specifications.

The **Financial assets** survey

- It contains 3819 forms among which 2846 variables and 43 files

The 1991 **Health** survey

- It contains 1516 forms among which 1033 variables, 111 specifications and 18 files.

The 1993 **Transport** survey

- It contains 1388 forms among which 993 variables and 12 classifications.

The 1995 "**Social Mobility and Integration**" survey has given birth to two distinct bases according to the concerned populations.

- The first one contains 1332 forms among which 821 variables, 37 specifications, 46 programs and 6 classifications.

- The second contains 1139 forms among which 656 variables, 108 programs and 7 classifications

29. In the area of business statistics:

The annual **Trade** survey

- It contains 3056 forms among which 726 variables

With the ISIS tool

- The **ISIS** base ( **I**nventory of the **S**tatistical **I**nformation **S**ystem)
- It contains 13541 forms which allow to make the census of 215 surveys and 9886 questions, 1033 publications and the definition of 565 concepts.

With the TEXO tool

- The documentary base of the **National Accounts**
- It contains 3472 forms which allow to store 107 publications and 700 abstracts.

For the bases which use the SURVEY tool, one will notice that they are, in a great proportion, composed of the definitions of the variables from a survey, linked of course to questions and to corresponding questionnaire(s). For the remaining, the documentation which makes up the bases is a function of everyone's needs. Some of them contain many specifications others very few. On the other hand, the source of the computer programs do not always appear, but above all an important lack has to be noticed when concepts are concerned.

#### The use of the DDS bases

30. When several **DDS** bases exist corresponding to successive versions of a survey, then the information contained in the first ones has been used for the more recent version at the stage of its design. For surveys for which there has been to this date only one documented version, two main functions have been used, the archives and the editing of the codebook with the help of the SAS interface (See **Annex 4**). For the surveys made in partnership, all the documentation has been provided to the different partners.

#### The Romanian achievements

31. In Romania the first concrete application of documenting a statistical operation has been done on the annual survey **AMIGO** (Romanian Employment survey) for the years 1994 and 1995. This base contains 238 forms among which 90 variables, 57 specifications, 3 classifications and 22 concepts. It gave rise to several uses. For example, all the documentation of the annual survey could be used for the management of the quarterly employment survey which has existed since 1996, both operations having many common

characteristics. The very base of the **AMIGO** survey has allowed to make an instruction guide for the interviewers, to edit the Codebook (See **Annex 5**) and the specifications for the programmers (checking conditions between variables). Moreover it allowed to rationalize the 1996 annual survey. Lastly, the base being particularly rich in concepts, it allowed to answer external demands coming with the data from the survey. This query has been made by the Forecast Institute, by the Institute for Quality of Life, by the Institute for Economic Research, the Ministries of Labour and of Finance.

32. The SURVEY tool has further been used to document macroeconomic indicators. To adapt to the situation of the "Survey model", indicators have been dealt with as variables. In this way their definitions could be documented. As far as it does not provide a complete answer to the problem dealt with the model has been "forced".

33. Thus, the SURVEY tool allows one link and only one between a variable and a classification. In the present case, it was necessary to link several classifications to one indicator. To make up for this drawback, the classifications have been introduced as comments to the texts of variables. This is of course not satisfactory since it is not a priori natural to search for classifications in the comments of the text of a variable.

#### The difficulties encountered

34. Two main difficulties have been encountered in the practical making of documentation bases. The first one is perfectly illustrated by the previous example. The constraining character of the data models which underlie the existing tools make their use tricky for operations the characteristics of which differ noticeably from those attached to the three data models. Even staying in the area of surveys, it is already true for business surveys. A fortiori, the phenomenon becomes still more obvious as soon as one gets interested in operations the source of which is not a classical survey : indexes, indicators, accounts, administrative files, etc... "To force" any of the three data models to adapt to the organization of the documentation of an operation can lead to inconsistencies in the contents of the bases, such as to deal with classifications as comments of a variable text. Going from one base to another, there is no longer a guarantee to have a consistent processing of information which secures an efficient search.

35. The second main difficulty is of a different nature. It is linked to the work load which the preparation of a documentation imposes in a context of relative shortage of human resources. This load which is derived from the working methods and from the technical characteristics of the current tools can discourage the designers of a statistical operation from documenting it, whereas the demand for technical documentation becomes more and more important.

36. This double established fact imposes a reflection as for the modifications to bring to this procedure in order that it be able to satisfy more the needs it must answer to. In the present document, we shall only tackle the part related to the tools, the one related to working methods being much more specific to each organization concerned by this question.

### III. THE PROSPECTS

37. The experience which has been accumulated in **INSEE** for several years and the more recent experience of **CNS** lead to a double conclusion.

38. On the one hand, the organization method for documentation as it is designed in the **DDS** is a satisfying answer to the problems encountered in this area. On the other hand, the operational aspect of **DDS** (the computerized tools) shows a number of inadequacies, the origin of which is both functional and linked to the fact that these tools, developed six years ago, need to adapt to the progress of computer technology. Thus they must develop in two directions. The first corresponds to an improvement of what exists. The second corresponds to taking into account needs which have not yet been satisfied, but which express themselves with more and more vivacity. One of these needs is linked to the necessity to generalize the documentation possibility of statistical operations ; the other is the recent development of dissemination techniques. Facing this situation, the decision has been taken, in total agreement between **INSEE** and **CNS** to proceed to a remodelling of the tools. It is being developed and should end up by the end of year 1996.

#### The improvement of what exists

39. Two points are to be considered:

- The constraints proper to the tools.
- The homogeneity between the tools.

The constraints proper to the tools

40. The main criticism expressed by the users on the existing tools deal with the following elements: the processing of text documents, the navigation in the bases, the management of the links, the ergonomic aspects.

The processing of text documents

41. In the present system, the text documents added to a **DDS** base must follow a triple constraint:

- Lines of a maximum of 72 characters.
- Loss of the initial enrichment of the documents. Any text recovery for printing from a **DDS** base requires that the enrichment be redone if one wishes to have a careful edition.
- No other type of document but text can be stored in a **DDS** base.

42. Thus the improvement consists in allowing that text can be stored in any base with no constraint of line length, that the enrichment of text can be retained and that one will have the possibility to introduce directly other types of documents (graphs, flowcharts...).

The navigation into the bases

43. In the present system, selecting a subset of forms may turn out rather complex in terms of handling of the computerized tools. This is the more disheartening as this type of query of a **DDS** base is of course basic. The remodelling will allow, with the help of graphic interfaces and of the navigation means offered by the **WINDOWS** technology, that the query mode of a base will be simplified and that text recovery will be easier.

#### The management of the links

44. One of the major interests of a documentary base built with a DBMS is to be able to introduce links between the forms. One will thus be able to specify precisely the description of each form as to navigate in the whole documentation with the help of these links. In the present tool system, adding a link requires manual data capture and memorization of the names of forms, as this can discourage the users to do so it may lead to impoverished bases. The improvement is to facilitate the introduction of the links, even to make it automated when possible.

#### The ergonomic aspects

45. Generally, the ergonomic aspects of the tools is linked to the technology of time when they have been developed. The improvement aims at using the possibilities offered by recent technology and userfriendliness of the computerized tools. On line help, integration of the tool to the working station of the statistician, more adapted communication with other tools for management operations.

#### Homogeneousness of the tools and communication between bases.

46. For the time being one does not have a unique tool whatever be the data model considered. This situation has the following consequences:

- Functionalities are not perfectly homogeneous between the different tools.
- From one tool to another, communication is not easy. For instance, one cannot copy directly information which can be found in a **SURVEY** base to feed **ISIS** and vice versa. This would be both a gain of time and of consistency. For instance, **ISIS** contains a rich information for concept definition and this information can be recovered easily in any particular **SURVEY** base.
- The different versions of the **SURVEY** tool are not consistent.
- Besides different versions of the word processor used, the **SAS** interface, from which one can in particular edit a code dictionary, exists only in the **MVS** version.

47. The improvement work thus consists in developing **a unique tool**, with homogeneous functionalities, whatever the data model used.

#### Extension of the tool capacities

Possibilities of adaptation of the model

48. The SURVEY tool is based on a data model explicitly linked to a type of statistical operation, the source of which is a survey and even more particularly a household survey. In the same way, ISIS and TEXO are based on a data model adapted to their source objects. All the cases which possibly are of interest to statisticians do not necessarily correspond to these fixed schemas. A statistical operation which is originated from administration, a continuous statistical operation such as the consumer price index, synthesis data such as the ones of the National Accounts, the processing of times series such as the Quarterly Accounts need a data model different from those which can be found in the present tools.

49. The documentation of time series can illustrate this. So, when one works with infra-annual series, an essential element is the statistical adjustments which are brought to raw data. It is a question of, for instance, seasonal adjustments ( CVS ), or working days adjustments ( CJO ). In that case, none of the types of form of the existing models is really adapted to taking into account of the information useful to understand the methods used in these processings. One must therefore imagine a new type of form, linked to the variable type which remains interesting. One could call it "processing" and it should be characterized by fields which would supply the following information.

Fields	Examples of field values
Nature of the processing	CVS, CVS-CJO, CJO-CVS
Frequency of the CVS processing	month, quarter, year
Frequency of the CJO processing	" "
Composition of the CVS processing	additive, multiplying
Composition of the CJO processing	" "
Type of adjustment for the CJO	working days, not working days
Beginning date of the calculation of CVS coefficients	
End date of the calculation of CVS coefficients	
Beginning date of the calculation of CJO coefficients	
End date of the calculation of CJO coefficients	
Beginning date to apply adjusting coefficients	
End date to apply adjusting coefficients	
Software used for CVS processing	
Software used for CJO processing	
Definition of aberrant points	

50. The texts of this type of form would allow to supply the detail of further methodological explanation : definition of aberrant points, choice of the CJO type of processing, choice of the order of the adjustments etc. What is true in this example is also true in many others. It is therefore

essential, if one wishes to be able to answer different queries, that the unique tool be adaptable to the different cases.

51. In this context, the organization of the bases should be modified. The role of the bases administrator is noticeably strengthened compared to the present situation. It will belong to his responsibility to implement, after a discussion with the designers of the statistical operation, the technical translation of the specificities of the chosen data model, in the initial framework of the organization method of the documentation. It will also be his role to make specificities consistent from one base to another. It is thus a question of protecting the possibility of exchange and recovery of information from one base to another, the same type of information appearing with the same description. If one looks, for instance, for instruction notes to the interviewers, these must appear, whatever the base, in the same type of form with identical field values. The role of the management clerks of the base is not modified. They have the same attributions as before.

52. The possibility of extension to different types of statistical operations can generate a need of relative confidentiality for some bases which can be "private". For this reason, the persons in charge of the query are divided into two categories: privileged or ordinary. In both cases, they may only carry out operations of query and/or recovery in a base.

- a privileged person in charge of the data base query is always declared to the administrator and has access to the information contained in any form.
- an ordinary query person has only access to "public" forms. Access to a so-called "private" database can only be given to persons declared to the administrator of the base.

#### Portability of the tool

53. The adaptability of the model is not the only way to remodel the tool to meet the need of its extended use. There is also its "portability". The present developments have been made from the same DBMS, chosen both by **INSEE** and **CNS**, which is **ORACLE**. This limits the possibilities of dissemination of the **DDS** instrument. As a consequence, the computerized tool must be provided with the capacity to work with other DBMS (Paradox, Sybase, Informix) in the same conditions as with **ORACLE**.

#### Dissemination of the metadata

54. The demand, which for a long time has remained latent, of metadata from the data users, becomes stronger and stronger, and it will probably have an important and quick development. It concerns firstly the disseminators. The situation which they judge ideal would be to have at their disposal an "on line help" they could easily query concerning the data themselves. As a consequence the tool must be provided with a new functionality, allowing to provide the automated generation of hypertext documents from the information

contained in a **DDS** base, the hypertext links copying the links defined in the base.

The expected benefits from extension of the tool capacities

55. From providing statisticians with a remodeled tool, a better rationalization of the whole production is expected, **whatever the nature of the statistical operation and whatever the place of establishment.**

- The designers will have an instrument to assist in the rationalization of the organization methods of a statistical operation (in the case of new operations), together with their transmission in the case when the keeping of this memory is a main component of the production conditions (continuous operations, limited, heavy and repeated operations).
- The management clerks will have easily accesible help to the specificities of progress of a statistical operation (transmission to management clerks or to different actors of a production chain of programs, of the documents necessary to their activity, either by direct access to the bases or through help files). This need is particularly important in the case when the statistical operation involves decentralised management methods.
- Other designers or management clerks exterior to the concerned statistical operation will be able to find the methodological information on statistical operation they are not in charge but which concerns them directly.
- The direct users of data produced by the statisticians will find the methodological information about the operations which originate the data they are using.
- The disseminators will be able to document their dissemination products when they will be asked to do so.
- The persons in charge of the archives will find the elements required to document what they wish to save.

**The means implemented**

56. They belong to two categories: technical and human.

The techniques used

57. The implementation of such an evolution would derive benefit of using two complementary techniques: **object-oriented programming** and **Client/Server architecture**:

- With the objectives of remodeling as for the improvement of ergonomic aspects of the management of a base, of the possibility of adaptation of the model and portability, it appears that **object programming** techniques are well adapted to the development of such an application. They allow to model the problems close to the real domain described and present great benefits as for reusability (easier maintenance, open possibilities of extension).
  
- Besides, the **Client/Server architecture** allows to answer the questions related to the flexibility of base establishment, of communication between bases and with other management tools, which allows the integration of the application to the work station of the statistician.

With this background, DELPHI is the development tool chosen. It reconciles the conceptual benefits of the object-oriented programming with the flexibility of the Client/Server approach concerning the establishment of data.

#### The manpower

58. This remodeling undertaking of the **DDS** tool is implemented on the basis of a close collaboration between both institutes (**INSEE** and **CNS**). Jointly and following the success of their previous experiences, they have decided to derive benefits from their expertise, both functional and technical, to work jointly for the development of this new product, the task distribution having been done according to each one's expertise.

#### **CONCLUSION**

59. The remodelling of the computerized tools of the **DDS** instrument is a necessary condition for the development of the setting-up of documentation bases of statistical operations. To the best, it is not a sufficient condition. As a matter of fact, the will to document statistical operations and the setting-up of the working methods necessary to it, are essential conditions to provide for this development. Now they belong to the domain of the organization of the concerned Institutions and no longer to the one of pure technique.

60. One can think however that this intention will develop in a near future, so the need to know the methodologies of compilation of statistical data increases. Consequently, a remodelled instrument should provide the producers of these data with a satisfying working tool.

---