

SIGMA

The bulletin of European statistics



No 2
March/April 1992 – New technologies ...

Statistical Office of the European Communities

C ONTENTS

Editorial

page 3

Publishers:

Office for Official Publications of the European Communities

Managing Editor: A. Chantraine,
Director of Dissemination, Eurostat

General coordination:
Eurostat – Public Relations

Design & editorial:
J. Drappier – C.R.E.A.M.

Graphics & lay-out:
Eurostat – Publications and database administration

Luxembourg: Office for Official Publications of the European Communities, 1992

Catalogue number:
CA-AB-92-002-EN-C

© ECSC-EEC-EAEC, Brussels,
Luxembourg, 1992

Reproduction is authorized, except for commercial purposes, provided the source is acknowledged.

Printed in Luxembourg

The points of view expressed in this publication are those of the authors. The opinions expressed in this document are not necessarily those of the Commission of the European Communities.

Cover photo: Astra 1C satellite
Courtesy of *Société européenne des satellites*, Betzdorf,
GD of Luxembourg

This publication may be obtained free of charge from

Eurostat
Statistical Office
of the European Communities
Bureau d'information
Bâtiment Jean Monnet
L-2920 Luxembourg

DIRECTIONS AND INDICATORS

Hi-tech statistics	4
Tests and projects at the INSEE	8

EUROPE AND STATISTICS

Remote sensing and statistics: a new partnership	13
Statistics at the dawn of the information age	16

MISSIONS, VISITS AND MEETINGS

19

INSIGHT INTO EUROSTAT

The many challenges of EDI	20
----------------------------	----

NEWS ITEMS

'New techniques and technologies for statistics' Bonn seminar, 24 to 26 February 1992	25
--	----

E DITORIAL

As the single European market takes shape so there is a growing need for comparable, reliable and rapidly available statistics in numerous sectors of the economy. This is therefore a fitting moment for Eurostat to take a front seat in the development of new statistical instruments.¹

The Doses programme² looks set to yield fruitful results, particularly in the form of applications which make use of artificial intelligence.

Businesses and government departments, researchers and teachers: all are concerned. Statistical information – usable in ‘real time’, easy to display, multimedia and multi-format – has taken on a strategic importance.

Tomorrow’s technologies are already being applied today. The exchange of computerized data (EDI) has led to increased competitiveness, and is obliging government departments to adapt swiftly. Remote sensing by satellite is yielding information of great value for agriculture and the environment.

The barriers separating the national statistical systems, at both European and world levels, are disappearing. The European statistical system will, like a ‘nervous system’, be broadly dispersed.

For businesses, application of the most advanced technologies will help to make the 21st century the ‘information century’.

Photis Nanopoulos

Director of Business Statistics

¹ See H. Christophersen – Preface to the brochure presenting the Doses programme (*Eurostat* publication, Catalogue No CA-55-89-762-EN-C).

² Doses: Development of statistical expert systems programme.

H I-TECH STATISTICS

An interview with Mr Yves Franchet

If we consider statistics from the technological point of view, we find that we are obliged to look at it from more or less every aspect. Statistical technology is there to be used by the entire Community statistical system; it has its role in every domain and in every policy which Eurostat pursues.

In the following pages it will be difficult for us to avoid mentioning some subjects more than once. We have done our best to arrange things in such a way that the impression of 'dejà-lu' which may be generated from time to time will in fact whet the reader's appetite for more of the same ...

WHAT NEW TECHNOLOGIES?

Statistics is a highly complex field: statisticians have recourse to a wide range of techniques, for each of which specific technologies permit the collection, processing and dissemination of quantitative data. As a Community body, Eurostat has the additional responsibility of harmonizing the data received from its partners, in order to lend comparability to data which must first of all be coherent.

The traditional techniques – some older than others – (e.g. collection techniques such as sampling, censuses, surveys and panels, standardization techniques, organization techniques and administration techniques) have been supplemented by the techniques nowadays derived from countless applications in a wide range of domains; these can be classified under three headings:

- telecommunications and remote transmission;
- satellite remote sensing; and very generally
- any other computing and office automation developments applicable within the field of statistics.

When we talk of these 'new technologies' in statistics, what should we be talking about: improving existing tools, developing new ones, or applying them to new domains?

'Not one of them on its own. The approach has to be "how can we use new technologies to make the statistician's job more efficient?"'

New skills

We have to look for the efficiency that new technologies can offer, in terms both of current activities and of new domains. New skilled ancillary jobs will appear. Look at satellites: they send us an enormous volume of data, practically in real time. Having to go and find the information is no longer the problem: our problem is processing it: it is arriving at such a speed that

we are obliged to bring in processing equipment fast enough to handle it; to build expert systems capable of managing gigantic quantities of data whose complexity is increasing all the time. And that, in turn, means further efforts to standardize, organize and coordinate: these new tools will only work with standardized, well-codified messages. That is what "new technologies" mean in statistics.'

MANY IMPLICATIONS

Eurostat is probably the Commission's biggest user of hi-tech resources. Does the Commission make allowances for that in the budget?

'The Commission's policy for developing new technologies reaches deep into the Member States, so that norms and standards can be promoted: this means lots of research projects.

But Directorates-General like Eurostat, heavily involved in management tasks, do sometimes have difficulty selling the idea that management systems need investment too.

All too often it is plain that resources are readily made available for research, but that Community financial incentives are lacking when the time comes to implement the results of that research. Although the Commission is making a considerable effort, management incentive projects do not receive the same sustained attention as research projects.'

Many 'new technologies' are developed with a view to economizing on manpower. Is there a risk of over-automation?

'Investment and training should go hand in hand. Initially, new technologies tend to bring an additional burden rather than any saving in human resources. It is only after some time that they are capable of allowing the same job to be done by fewer people, or more work to be done by the same number. Our own development to date has been in this latter direction:

with the same staff we are now producing better work, different work.

Is there a risk of over-automation? I rather doubt it. In an administrative system like ours, where the idea of investing today for savings in the future is scarcely an automatic reaction, we are constrained by budgets, and our progress is therefore necessarily gradual.

Training

We must train people, of course. No one trains their staff enough. They think it's all too simple. Office systems are used far below their true capacity because the users haven't been trained properly.

Another classic problem with new technologies is the generation gap ... the older staff tend to be in the senior jobs, and that creates problems which have to be solved by changing the way work is organized. I have to say that in Eurostat this aspect seems to be working out fairly satisfactorily.

The cost/efficiency ratio

The job of Community statistics is to produce information which previously existed nowhere else. We have to produce it in the right form, for a widening range of domains, and inevitably we are facing a time of rising costs. A policy of systematic cost-cutting is not realistic.

So our aim is that the gain in efficiency should be equal to, if not greater than, the growth in costs. And in a world where efficiency and high technology go hand in hand, we also have to be on the look-out for the often overlooked costs of things left undone, investments made too late, and so on.

The Commission does not, *strictu sensu*, have an investment budget, a maintenance budget or a depreciation budget: any item bought is deemed consumed, and large-scale investments, which may produce effective savings only after some time, are not just difficult to programme: it is also difficult to put an immediate figure on their impact on efficiency.

Lightening the burden on the data-suppliers

The people who supply statistical data – particularly businesses whose data management is already largely computerized – are

upset at the price it costs them to supply the basic data, the more so since they may themselves not yet be aware just how much a broad-based and efficiently-managed statistical system can offer them in return.

The application of advanced techniques to statistics should enable us to reduce the burden on businesses.

It is up to statisticians to make their input – and make it now: any later will be too late – so that the statistics which already form part of the mass of data needed by national and Community administrations can be extracted automatically as a by-product of business management, either directly or via the competent trade or professional associations. Care must naturally be taken to ensure that the data are treated with the confidentiality they require.

The statisticians' input

Statistics has to find its place among a cluster of new jobs we are likely to see developing over the next few years: the jobs at the interface between business management and the need for administrative data. We must do away with paper and inject a degree of statistical awareness into the interfaces and applications software used by accountants and auditors, by the broking organizations and agents, so that gradually we begin to receive data direct, down the line, on direct computer links ...'.

PRODUCTS ... PROJECTS ... PLANS ...

CD-ROM

The new technologies are generating new products in data-dissemination – CD-ROM for example. Is Eurostat ready to go into this market?

Of course. Comext on CD-ROM is going very well indeed (See *Sigma* No 2, November/December 1991, p. 23). Hardware costs are falling steeply, and the number of users will rise fast. The beauty of CD-ROM is that we can download our updated databases regularly and distribute them on a medium which costs practically nothing and is child's play to use.

We have to be pragmatic in our approach to the market: launch a

product, watch the reaction, adapt: there are major changes waiting for us.'

Does adding new media like these to the benefits of more intensive use of remote transmission mean that we are moving towards even greater decentralization in the production of statistics – more subsidiarity, with Eurostat ultimately playing no more than a coordinating role?

'The growth in data transmission does mean that Member States are standardizing the data they send to a greater extent than ever. Should we go further? It's not inconceivable that specific statistics – tourism, for example – might be centralized somewhere other than Eurostat, in a European Tourism Observatory in one of the Member States, perhaps. Data which are anyway publicly available, at least.

Artificial intelligence

Our own statistical research programme within the major Community programmes on artificial intelligence (Doses in particular) might be summed up as an expert system whose purpose is to improve the efficiency of statistical work. Our aim is to create intelligent interfaces enabling us to process data faster, organize nomenclatures, accelerate data merging, test data, and so on.

This research allows us to link together the concerns of a number of research institutes and organizations in the Member States, which without this coordination would only be able to work together with difficulty, if at all.

Tangible results

The Doses programme was launched in 1989, and its results with direct statistical applications have just been evaluated (see "New techniques and technologies for statistics" under the heading "News items" in this issue).

Blaise is a typical example of the high technology we feel we should be promoting: a portable computer system for household surveys, developed by our colleagues in the Netherlands. We have made efforts to adapt it for the Community, and are financing its maintenance and development. It is already being used in many

Member States, and the East European countries are now also interested.'

The greater use of new technologies is likely to take us very close to the borderline between the administrative survey and the survey for tax purposes. Is there not a danger in that?

'We take care that individual data are protected. There is no need to be haunted by Big Brother: we can and do ensure that confidentiality is not breached, and we are improving the techniques all the time.'

In Denmark any "administrative operation" by a Danish citizen is encoded, all the files are merged under that same code and the results go to the Danish Statistical Office, which then knows that the occupier of such-and-such a house is called X; they know what his job is, what he earns, how many children he has, and so on. This doesn't prevent democracy from thriving in Denmark.'

THE END OF THE PAPER ERA

Customs documents at the Community's internal frontiers will disappear shortly, to be replaced by a new system of data collection. Does this mean that statistics could have some role in measuring how far rules on competition are respected?

'Customs documents have been a traditional source of reliable data for generations. They are on their way out, and we are replacing them with Intrastat. Basically, Intrastat is a system using the business statistics derived from the document businesses have to complete to pay VAT.'

Critical choice

These data will not be of the same quality as the customs data previously used, and precisely because the intention was to lighten the burden on businesses, small firms will be exempted from making a statistical return. Broadly speaking, we shall be getting 95% by "surveying" 20 to 30% of companies.

This will represent a major simplification, albeit at the risk of losing a degree of reliability, but this system will nevertheless provide us with a clear picture of the main movements. The very substantial

savings which should result were a decisive factor in the choice made by the Council and by Member States.

Keeping an eye on competition

People whose job it is to keep an eye on competition – and that means businesses themselves, first of all – will still be able to identify anomalies, shifts and changes in the market. It will be for the competition experts to seek out their causes. Intrastat will take its place amongst the sources of intra-Community statistics, and will be used by the competition departments to ensure that the game is played according to the rules. Just as customs data are used today.'

CHANGING REQUIREMENTS OF BUSINESSES

Business demand for statistics used to be fairly elementary (construction, rents, wages and salaries, etc.) but is now becoming more specialized, as though businesses were discovering statistics as a previously under-used management and development tool. But while business demand is increasingly for tailor-made statistics, the supply seems increasingly to consist of off-the-peg, standardized products. How can supply be matched to demand?

'It is certainly true that in the past business statistics were principally – perhaps excessively – produced with the needs of government in mind. Businesses themselves had no part in it, and received no feedback.'

We are trying to put that right. For the first time business associations, the European industrial branch federations and European employers' associations have been closely involved in revising the nomenclatures of products and activities. The results of this work are likely to have a major impact because they meet the needs of businesses as well as those of governments.

A major collective effort

We have seen, too, that European business federations are ready to join together in European Economic Interest Groupings (EEIGs) to work together, with us, to improve the quality and particularly

the freshness of statistical data. What we are trying to do is to establish a closer relationship between the annual survey data and the fresh data given to us by the members of a particular sector, so that we can draw meaningful conclusions from the fresh data as quickly as possible. If the production of official statistics to meet the needs of governments is structured in such a way that it also meets the needs of businesses, the result will be closer cooperation, and we shall be able to move towards a system capable of taking up in a single sweep all the information that is needed by anyone.

That's why it is important that we bring the suppliers and users of statistical data into our plans. That is how you match supply and demand. What an EEIG has to do is start with a reasonably compatible set of data and then develop from it other statistical uses, perhaps other, more detailed forms of survey, which then meet the needs of businesses. The result will be a better match between supply and demand on the one hand and greater compatibility between governments and businesses on the other.'

Technology is synonymous with standardization. That means developing nomenclatures which can satisfy the requirements of both governments and surveys. How can that be coordinated?

A server for nomenclatures

'That's right, there is a relationship between these nomenclatures, and it's essential to be able to move between them in order to see the full picture. And, of course, they must be available to everybody involved. Hence our current examination of how we can set up a nomenclatures server – a computer application capable of giving all Member States, and all our partners in every Member State, an organized and properly coordinated means of access to the entire nomenclatures system.'

An application such as this will enable us to follow through all the implications of any change made at any point and at any time. We are working on this project jointly with the French statistical service, and it should be ready for application within a few months.

Other applications?

We're thinking further into the future already. Satellite images, for example, have to be interpreted. We are working with the Joint Research Centre in Ispra on what one might call an "interpretation server", capable of coping in real time with the flood of data which is constantly being beamed down by the satellites.

This sector, in which we are working closely with the Directorate-General responsible for research programmes, is assured of a very important future.'

A EUROPEAN ACCOUNTING CHART

In business statistics, a European accounting chart would without any doubt be the best guarantee of perfect comparability of data. The European Statistical System clearly needs a minimum standardized accounting chart to be adopted as soon as possible. How is the next statistical programme going to tackle this?

'Unfortunately there are two contrasting approaches to accounting within the Community: the French and the Anglo-Saxon.

The Fourth Directive deals with this. An attempted revision last year tried to make the point that statistical data would be much improved if some common ground could be found between the two approaches, but Member States unfortunately decided to take no action.

Personally, I am convinced that this is very important: vast amounts of information in accounts could be used, from computerized accounting centres. But if the accounting charts do not keep to the same definitions, we are left with no more than pious hopes, even after the recommendations of the International Accountancy Association.

I can't see the problem being settled for some years. It's a pity.'

AGRICULTURE AND STATISTICS

What can new technologies offer to support the reconstruction of the common agricultural policy?

'Reconstructing the CAP is first of all a political aim: transforming support for the product into support for the producer.

The challenge

Supporting the product was beneficial in the 1960s. More recently it has led to excess production and higher prices for the consumer. It has also had negative effects on the GATT negotiations.

The challenge is to set up a system of support for the producer which is neither an encouragement to inefficiency nor an excuse for failing to move with the times.

The role of statistics

- We have set up systems for measuring the farmer's global income: more and more holdings generate not only agricultural income but also income from craft or even industrial activities. So we have turned our attention more to the production unit.
- While that has little to do with new technologies in itself, remote sensing by satellite, on the other hand, may allow us to measure more accurately certain product premiums and, even after they have been abolished, to assess and pinpoint more accurately the overall effect of the CAP.'

GEOGRAPHICAL INFORMATION SYSTEMS

'If we look beyond agriculture, the need for machine-readable geographical data is becoming much clearer in countless domains.

One essential question

Eurostat is a Statistical Office. Should it also become a Geographical Data Office?

It's beginning to seem that it should. Indeed it's virtually inevitable the moment one talks in terms of the "environment". And if geographical classification systems don't match the systems used for regional statistics, for example, we shall have the problem of overlapping nomenclatures

and incorrect data. We have spent a year thinking about it, and now we have decided to expand into this area.

And from there ...

A good geographical information system is the only way of handling the data we get from satellite detection.

And it is also time that we introduced this geographical coding into a number of administrative surveys.

At the same time, with the development of the European Economic Area, we must be able to measure what there is in any region, pinpoint it easily in terms of geography and cartography, and match the right aid to the right place with a minimum of delay.

MUCH MORE THAN STATISTICS

There will be spin-off, much of it unexpected, from the faster transmission speeds, and greater accuracy in both the measurement and interpretation of satellite images.

With satellite images, for example, we shall know in April what state the land under cereals is in during April. For the present, we only have the information in June, by which time it is too late to deal with crop diseases or limit the losses due to excess rain or drought.

If we apply that system to the countries of Central and Eastern Europe, we could improve grain production in the Ukraine or Siberia ... Current estimates suggest that poor distribution, or the wrong treatment at the wrong time, mean that those regions lose 25% of their harvest. That's twice the volume of their cereal imports from the USA!

There are countless examples outside agriculture, too, all showing that these new technologies are going to have some quite extraordinary consequences. Well beyond the field of statistics ...'

¹ Interview with Mr Yves Franchet, Director-General, Eurostat, 14 January 1992.

T

ESTS AND PROJECTS AT THE INSEE

Interview with G. Raulin and F. Vennat

***Sigma*, the Bulletin of European Statistics, is a means of communication aimed more particularly at all those involved in the creation and enlargement of the European statistical system.**

It embraces the Commission and the Community Institutions, Parliament, the Economic and Social Committee, the Council, international organizations, various public and private users, as well as the NSIs.

It is in this context that the INSEE appears here and that all the members of *Sigma*'s normal audience will subsequently be invited to express their views in this column.

G rard Raulin

A graduate of the Paris  cole polytechnique and trained as a statistician at the ENSAE, he has held various posts at the INSEE over the past 24 years. He is now Head of the Projects Department and is responsible for new developments and new applications.

Francis Vennat

With a master's degree in informatics and an ordinary degree in sociology and trained in economic statistics at the CEPE (Economic Programme Study Centre), he is Head of the 'Internal projects and studies' Division in the Informatics Department, which deals with infrastructure, technical projects and the maintenance of existing applications.

Since the lines that follow stem from a largely improvised interview, the ideas presented do not necessarily follow a logical pattern. 'Because the reader is not paid to read it, a review article cannot be written like an official note, and still less like an essay ... One has to lose certain habits and acquire others ... What is required is not to shorten the amount of information but to shorten the path from the author to the reader'.¹

FROM THE PRESENT TO THE FUTURE

A rapid overview of the three levels of activity of an NSI will show the technology in place and its likely evolution.

Collection

'It is perhaps at this level that, compared with other countries, progress in France has been the most difficult, except in the case of enterprises. Although for surveys proper the good old paper questionnaires are still being used, there are changes under way in annual salary declarations, for example, which in a very few years will bring us to the stage where everything will be done by file transmission, or even EDI.

We will deal elsewhere at some length with certain recently established services (Tedeco) for exchanging data with local authorities.

This year also sees the introduction of a procedure for collecting data on households using a laptop computer. The system could ultimately lead to the organization of smaller-scale surveys, reduce survey times and simplify the intermediate coding and input phases.

We use the Blaise system, which allows us to maintain excellent contacts with our colleagues in the Netherlands, and we cannot but admire their performance given that it is extremely difficult for an NSI to provide the back-up for and dissemination of such a product.⁷

Processing

Of the conventional or new methods currently being used, we consider it useful to concentrate on areas involving automatic coding procedures.

Dissemination

'This level depends to a very large degree on the general efficiency of the system. Technological progress has certainly resulted in advantages as regards:

- access: computerization, databases and telecommunications provide greatly improved access to all the data managed;
- storage: bulk storage is possible, with direct searches on computer media, providing greater speed of access;
- response modulation: we are moving towards systems that will allow access to sets of data spread over different logical media, in files of different formats and different classifications, using telecommunications techniques that will link central bases and microcomputers, with much improved presentation techniques than simple conventional tables, such as chart and map outputs.⁷

INNOVATIVE APPLICATIONS

We have chosen two innovative applications, among the many that could be mentioned. The first has been designed and developed since 1979-80.

Quid

Automatic coding

'This method of automatic coding was first used in the census in the French overseas departments.

Two other major INSEE surveys also employ Quid at present:

- the annual returns of social data (DADS);
- the employment survey.

The basic approach of Quid (abbreviation for identification questionnaire) is to create a very large

database consisting of typical categories of respondents, together with corresponding codes allocated by an expert.

Using a learning file, the system creates a question tree structure by interrogating the headings to identify a code.

Quid uses information theory to construct a tree structure of minimal average length in order to reduce as far as possible the time needed to identify the code.

The next version of the system, Quid2, will be based on tables in the learning file, which it will no longer manage heading by heading but by groups of headings from the same table. New correction and enhancement tools for table management will be installed.²

Applications

'This tool is in a semi-stabilized state and is being developed gradually. However, it has not yet been used for coding purposes in the population census in France. With the current survey documents, it was thought that the rejection rates would be too high and would require excessively complex processing.'⁷

Tedeco: EDI

Today, all economic operators generate ever more information that can be transmitted from one partner to another. The need for maximum security, transparency and speed in the transfer of information and the widespread use of computers by businesses make electronic data interchange (EDI) the most suitable system for the purpose. EDI avoids the need for manual re-input of information and facilitates exchange by collecting data in a directly utilizable computer form. Transfer from application to application is undertaken automatically in a few minutes in complete security. Tedeco is therefore not only a valuable and effective tool but also an essential asset for companies, businesses and organizations wishing to increase their productivity and the quality of their information exchanges.'³

A bridge between the public and private sectors

'Initially designed to facilitate and develop electronic data interchange between the Ministry of Finance and its external corre-

spondents, the system has been used experimentally since July 1990 between the public accounts sector (departmental Treasuries), the INSEE and local authorities (towns, urban districts, county councils, inter-communal associations and regions), and between the customs services and business ... The service will henceforth be accessible without restriction to all who wish to exchange data. To that end, the Ministry of Economic Affairs should create ... a Tedeco working party for the promotion and development of this new service

Tedeco is an open and standardized value-added service for the electronic transfer of data, safeguarded by an EDI protocol, between multiple partners

The software contains value-added functions such as data compression, segmentation of large files, and an opening notice applicable throughout in accordance with the recommendations of the P-EDI protocol. It will incorporate all the Edifact functions

The major users of Tedeco are municipalities of over 30 000 inhabitants, all district and regional councils and regional and local authorities that need to exchange information electronically with the Treasury and the INSEE, including subsidized-housing offices, regional hospital centres and chambers of commerce and industry.'⁴

And businesses?

'Businesses are the next stage. In any case, local authorities are required to forward to us everything that Tedeco can now supply and this reduces the burden on them. The system is only of interest to businesses if the information already exists in a standard, harmonized format from which it can be extracted easily and forwarded by EDI. That means that the business must already employ EDI for all its activities.'⁷

MAKING FIGURES SPEAK

What do figures mean without comments? What use are data without meta-data?

Automatic generation of comments on economic trends.

SAM

'Electronic data processing allows generation of statistical data at

low cost, but the resulting deluge of figures rapidly submerges any economist responsible for analysing them. Statistical analysis is of great help, but very different from human analysis. The latter always places the bare figures in a context; for example, when analysing time series the economist concentrates on the near present and does not seek the "averages" generally employed by statistical processes.

Human analysis is therefore always essential, but unfortunately has two defects:

- the often high cost of human resources;
- the inevitable inability of the human brain to select and highlight "important" information in a large mass of data. 200 monthly statistical series over two years, for example, represent approximately 4 500 figures to be analysed.

EDP, and more particularly artificial-intelligence techniques, fortunately provide more elegant solutions. The SAM system presented below is one such solution. It operates as follows: working around a numerical database, SAM has a database which describes the semantics of the application and supplies dictionaries. SAM itself is only a program, previously written in LISP and now available in C. Using the information in the two databases, SAM effects an analysis and drafts a text relating to that analysis in a manner similar to that of an economist. SAM also has a learning module which, by means of a set of appropriate questionnaires, allows it to create or amend an expert database. SAM can also draft a text describing the expert database which allows the user to compare his own knowledge and the factors on which he bases his judgements.

Although SAM is still at the research stage, it is aimed at improving efficiency and is destined to become a marketable product

SAM is currently being tested on different applications and is undergoing further development for the addition of narrative structures, focus management and documentation. In the short term it could be used at the INSEE to analyse individual replies to economic-trend surveys by providing a specific commentary on each enterprise that has participated in a survey.'⁵

Producing models of the economy and economists!

LCE (economic trend analysis software)

- Models of the economy:

'While it is true to say that mathematical analysis of economic theory has now been almost fully completed, the reliability of econometric forecasts has hardly made any progress for years. Paradoxically, therefore, the role of the economist in interpreting or fine-tuning the results of economic equations has never been as important. If mathematical modelling of economic theory (theories?) has taken some 25 years, perhaps the next 25 years should be devoted to modelling economists' thought processes.

- Models of the economist?

Producing models of the economist is more complex and quite different from producing models of the economy.

In the latter case, it is a question of defining a formal system of equations that mainly describes generally recognized behaviour of the components of the economic system. In the former case, it is a question of producing a model of the reasoning and judgement processes of a human being with a profound knowledge of the subject in question. Put simply, it is a question of simulating the intelligence and knowledge that enable us, by means of little-known processes, to produce from figures and other non-numerical data a coherent analysis of the state of an economic system ...

The LCE software allows all data types and structures to be presented and manipulated numerically and symbolically, in the traditional sense and in the context of artificial intelligence with a pronounced statistical leaning.⁶

FROM IDEAS TO PROJECTS

Cooperation between Eurostat and the NSIs has led, at the INSEE, to the creation of computer links and telecommunications procedures, particularly under the Caddia programme and projects such as Stadium and Sta-

tel. This year it will require a choice to be made on the use of the Tedeco service.

Under the continuation of the DSIS programme, the NSIs are involved in projects on distributed processing with decentralized databases.

The major revisions of classifications will have important implications for users. In order to facilitate the switch to the new classifications for both public and private users and ensure their optimum use in the future, the INSEE, in cooperation with Eurostat, is developing a server system project.

A server for classifications

The starting point for a wider extension, the project has two components:

Logical server

This tool is intended for professional users of classifications. It will allow direct access to the various classifications and parts thereof by means of codes. Sound knowledge of this tool is obviously required for its use.

Linguistic server

'The most ambitious part of the project will be the use of a linguistic server as a tool for accessing elements of the different classifications using texts written in natural language.

Although it goes far beyond the alpha paper index, even installed on a database with a rapid search facility, the server cannot be considered an automatic coding tool. It is on the periphery of the expert system and accumulates knowledge on vocabulary equivalence, and even has the ability in some cases to analyse sentence structure.

The server does not attempt to describe possible meanings but accumulates knowledge on the vocabulary used for a sub-set of servers (and, if possible, users) from which a motor proposes codes suggested by the sentence or phrase.

Its design makes it possible to utilize a large part of the work done in one language when one wishes to do the same work in another language. If, when switching from one language to another, some work on vocabulary has to be redone, a certain degree of independence between the vocabu-

laries and the semantic structures and correspondence between the meanings and the codes in the classification obviate the need to rebuild the whole.⁷

Businesses are affected

The reactions of all professional and occasional users of current classifications will need to change.

Businesses are affected in at least two ways:

- firstly, by their own code, whose correct identification has important repercussions not only in terms of information but also as regards the application and coverage of numerous regulations;
- secondly, something that hardly existed in 1973 (when the previous classification was developed), many businesses use the classification codes themselves to classify clients and suppliers in their files, compile files of prospective clients or undertake economic analyses. They use the classifications not only as their own references but also to compare the results of internal studies with external results provided by the statistical system or intermediate bodies such as trade associations, chambers of commerce, etc.

Modelling aid

'Recently, a modelling-aid system has been developed, based initially on an expert system, in order to help the modeller solve the problem of "tuning" the model.

Thanks to artificial intelligence

Briefly, a model is a set of equations for a given point in time and with validated reliability that is used to represent the economy and prepare forecasts. The model contains external and internal variables calculated for the periods for which forecasts are being made. What is also needed is a series of small adjustment variables.

Faced with systems that may comprise thousands of equations, the task of the modeller at this stage is to determine these small adjustment variables. He works by taking individual variables and examining the consequences one by one, a painstaking and prodigious task.

The proposed system will greatly simplify the work of modellers by

allowing the choices to be narrowed down more easily through a series of logical deductions.⁸

Electronic document management

'Another project in preparation is one aimed at replacing, in a number of domains, the onerous handling of huge amounts of paper by electronic document archive and management systems.

Confidentiality

We are holding discussions with the national informatics and freedom committee, which is to rule on the question of whether or not copies of documents on magnetic media should be considered as named files. Seen in a totally different light by businesses, the problem of confidentiality is of very great concern to households and families. Protection of privacy in general remains a very sensitive area.⁷

Disagreement among experts

How far should one go with artificial intelligence and expert systems in the search for statistical applications for businesses and individuals?

'We find that all those who, like us, have adopted this approach are encountering the same hurdles: difficulties in collecting and classifying data, availability of experts, choice of suitable subjects/applications, etc.

Furthermore, the specialists in these fields and in related disciplines do not always agree among themselves on how to define the areas of their expertise ...'⁷

MANY IMPLICATIONS

Today there is a widespread need to cut operating costs. If properly implemented, automation, rationalization and new technologies should normally lead to a product that is more stable overall.

Central systems

Coding: a changing job

Good human coders solve more difficult cases than machines, but they are more subject to error through omission, distraction or negligence. The coder's job will change but there will always be a demand for expert coders who ulti-

mately will deal only with solving marginal cases so as to improve the files.

Corruption of the job?

Does the statistician need to become a cartographer? And with the advent of desktop publishing has he not already become a compositor?

Rather than a corruption this represents a successful integration with the job of the statistician. Integration is possible and if measures to stimulate such integration are taken upstream, increased profitability rapidly results for all parties. In general, all successful geographical or cartographical projects undertaken to date by statisticians have owed their success to cooperation with specialists in those fields and have not been a matter of improvisation.⁷

Local networks

'Should local networks be classed under "new technologies"? The answer is probably yes, since such networks have been fully developed and meet the requirements of independence between central units and recognized decentralized establishments.

- At the purely technical level
In our view, the technology has stabilized but will undergo further very rapid change in an uncertain environment in some sectors, with alliances between major manufacturers and large software houses.
- Impact on statisticians
One should expect in the relatively near future a change in relations between computer managers and computer users. What will follow will be the establishment of independence at a decentralized level, with local responsibilities and the possibility of allowing local matters to be managed locally.
This step should be managed by providing general technical and organizational operating rules and be accompanied by a determination to avoid conflict between management and independence.⁸

Informatics is changing

'Increasingly, solutions to informatics problems will not be only at the technical level: more coordination will be needed, using tools for

designing information systems, applications developments and more strategic objectives.⁸

GENERAL DESIRE FOR COORDINATION

In all these areas, the Commission's role is to coordinate initiatives. Clarification would appear to be needed in the area of remote transmission (from NSI to NSI or between NSIs and the Commission).

Status list

'To our knowledge, the Commission has not yet produced a document showing the technologies, standards, etc. that exist in the various Member States. One would like to know, for example, where X25 can be used today and where it cannot be used; when it will be possible to use it and where and when a switch to X400 is planned, etc.

Better knowledge of the state of progress of everyone would make it easier for the various partners to prepare proposals applicable to all.

Such a document would provide us with a valuable reference for our internal choices, which must be made in the light of European integration.

Away from jargon

In the management of its projects and in its regular monitoring and evaluation plans, the Commission should seek to adhere to standard notions and employ fewer "in-house concepts" that we sometimes have difficulty in understanding.⁸

Active partnership!

'Partnership is increasingly essential at a time when many national projects are being undertaken. Such projects require closer coordination than in the past. The NSIs see partnership in a positive light, provided that it allows them to play an active and not a waiting role. Coordination will be successful only if it results from the combination of proposals from the different parties.'⁷

LEDA 2: time to relaunch cooperation?

'We are working on LEDA 2, a new generation of statistical software (the first LEDA, used at the

INSEE for the past 20 years or so, is employed for adjustments, coding and tabulation).

Are the other NSIs of the same opinion? Was this the opportunity to relaunch a cooperation project?

We are convinced that, when dealing with well-defined subjects, such cooperative effort can produce easily implementable results and benefits for all parties.⁸

¹ INSEE, *Courrier des Statistiques*, No 54, October 1990, p. 65.

² INSEE, Projects Department, L. Viglino – Communication on automatic coding, ISI 91, Cairo.

³ Preface to the Tedeco presentation brochure.

⁴ EDI Letter No 15, 31 October 1991.

⁵ 'Une synthèse économique rédigée par un économiste de synthèse', by Jean Louis Roos, INSEE, 13 December 1990.

⁶ INSEE, Informatics Department, J. L. Roos, 18 March 1988.

⁷ Gérard Raulin, 11 February 1992.

⁸ Francis Vennat, 11 February 1992.

REMOTE SENSING AND STATISTICS: A NEW PARTNERSHIP

Interview with Mr David Heath

'The impact of the availability of geo-referenced data on the traditional statistical working procedures has increased considerably in the past years. In particular, data from earth observation by means of remote sensing are a good example of such information. Today, enormous quantities of information about, *inter alia*, agriculture, environment, climate and land use become available with great regularity. The statistical information system is undoubtedly influenced by these new "geo-referenced" data, which are particularly relevant to various new areas of public and political concern.'¹

After studying philosophy, politics and economics at Oxford, David Heath qualified as a statistician and began his professional life in the UK Department of Health. He worked on the statistical and economic aspects of the agri-foodstuffs sector first of all in private industry and then in a semi-governmental body.

He arrived at the Commission when the United Kingdom joined the Community, and held various posts in Eurostat before becoming Head of the Agricultural, Fisheries and Environmental Statistics Directorate.

Remote sensing systems provide a wide range of geo-referenced data which can be applied, in particular, to geographical information systems (GIS). The GIS applications, which we will merely mention here, are to be described at greater length elsewhere.

A TECHNOLOGY OF THE FUTURE APPLIED TO THE PRESENT

How can information be obtained on physical phenomena without the actual presence of the observer?

Well-tested methods

'The development of remote sensing was to a large extent driven by military requirements and dates back to the first radars, aerial photographs and weather probes. The use of remote sensing for civilian purposes, while still relying to a large extent on photogrammetry, nowadays also makes use of satellite detection to observe phenomena on the earth.

The collection of information about the earth, the sea or the atmosphere involves the same principle. This consists of analysing and interpreting the waves and colours received by the sensors (active sensors that recover signals emitted by themselves, or passive sensors which recover signals, in particular light signals, emitted by the observed object and reflected after coming from sources other than the satellite).

Although apparently simple, the observation technique is in fact very complex. The angle of incidence of the satellite beam varies constantly. The satellite sways and turns on its orbit. Even if it is geostationary, the sun's rays can change angle at any moment. All observations therefore have to be constantly radiometrically and geometrically corrected to make them comparable.

Obviously all this is beyond the competence of statisticians. Such work should have already been carried out before statisticians receive the observations.'²

Large-scale projects

Routine use of remote sensing will soon be possible in statistics following, in particular, two big Commission projects.

The Corine project

'Developed between 1985 and 1990, the Corine system is designed to provide essential assistance for the European Environment Agency ... It makes it possible to collect and present, in a form which allows comparisons at EC level, very precise information on atmospheric pollution, soil erosion, water quality, the climate and nuclear power plants. ... It will be capable of showing, in table, list or map form, the state of our environment and the dangers threatening it.'³

The MARS project (Monitoring agriculture by remote sensing)

'Since the programme involves the application of remote sensing to agricultural statistics, it obviously concerns Eurostat and DG VI (Agriculture); the Institute for Remote Sensing Applications, which forms part of the Joint Research Centre, is involved in its implementation.

Eurostat must now get ready to receive the programme's first results, which are almost ready for use.

The cost/benefit aspect

Remote sensing cannot, of course, provide any information about the financial aspects of the problems observed, individual intentions, etc. It will be useful in areas where it will enable the observation of phenomena that lie beyond the scope of traditional methods. The Joint Research

Centre has proposed other very ambitious projects to the Commission, namely:

- 'Ocean': the observation of the seas;
- 'Trees': the study of tropical forests.

These are very high-cost projects, since the whole process has to be paid for, from the beam emitted by the satellite to the final product arriving on the statistician's or scientist's desk.

Statisticians require multi-purpose tools. Remote sensing will achieve its full potential when these products are available directly on the market, in the form of standard, commonplace images, which will be included in the information systems not only of statisticians but also those for the use of local authorities, municipalities, etc.

The cost, which today is much greater than the benefit, will drop appreciably and the trend will then rapidly begin to swing the other way.²

THE INTEREST FOR STATISTICIANS

It should first of all be stressed that remote sensing should not be made to compete with the old traditional statistical tools: in the statistician's 'arsenal', it is simply a new, powerful weapon that will be particularly useful in certain areas and should be integrated into the existing statistical system.

'Why does this new technology offer such interesting prospects for the statistician?

Objectivity

This is the great attraction of remote sensing, which makes it very relevant, for instance in the fight against fraud. The EAGGF Directorate of DG VI sets great store by this aspect.

Such a system does not depend on replies from a mass of separate respondents who may, in certain circumstances, have an individual or collective interest in slanting their replies in certain directions.

Comparability

Data from remote sensing will always have a constant approach, since they originate from a single processing chain. Using figures calibrated by this method, it will be

possible to adjust other figures produced by the Member States in order to make them more comparable.

A transnational vision

Remote sensing can easily cover phenomena which are larger than a single country, for instance all the data concerning Rhine pollution, the state of forests which cross national boundaries, the immense sea fisheries, etc. Remote sensing is very suitable for such purposes.

Constantly fresh data

The satellite gives a new image with each revolution it makes. Crop development can be monitored very regularly. The overall consequences of natural calamities, such as droughts and floods, can be readily and quickly assessed; this is also true of ecological disasters.

The mapping of phenomena patterns

This type of observation enables phenomena which are invisible at ground level to be detected and precisely located, for instance plant disease, pollution, water stress, geological features, archaeological sites, etc.

Observation of the environment

Remote sensing is particularly suitable for gathering information for the future European Environment Agency.

No need for on-the-spot observers

For certain applications, it is no longer necessary to send an individual to a particular site or to seek information from somebody already there. Places that are inaccessible to man are readily accessible to the satellite's "eye".

No more form-filling

Private individuals and companies no longer need to provide information. The response burden cannot be used as an argument against surveys carried out in this way. This makes collection much more flexible.

Geocoded information

The possibility of geocoding is becoming increasingly important in many areas of work. A precise geographical cross-reference very often has to be given to elementary statistical data. Modern com-

puter technology can then allow regional aggregates to be built up in varying ways to suit the particular problem being examined. In this way data are no longer limited to standard administrative regions.

This is particularly true for environmental problems, which do not stop at borders ...⁴

Many other Commission policies could also benefit from geocoding both those which are already well developed and those which are still in the early stages; for instance, there could be geocoding of regional data, town planning, etc.

ADVICE TO BUSINESSMEN

'Although remote sensing has developed less rapidly than some initially expected, it is now making very clear progress. Businessmen are now needed who are prepared to take the risk of putting products on the market ...'

A clear complement to traditional methods

Remote sensing will become an additional statistical tool and will be used:

- to assist collection, by considerably improving sampling procedures, calibration, comparability, etc;
- as an additional information source in the many areas which are difficult to survey by traditional means;
- as a replacement for old methods, etc.²

The statistician's contribution

Statisticians also have a lot to contribute to remote sensing, which has so far been too 'technology driven'.

'Those working in remote sensing must endeavour to cooperate with us. We are a big, well-organized structured client and have large dissemination systems which could be used to help remote sensing achieve its full potential.

Multi-purpose information systems are needed in remote sensing. It should be possible to use the same images several times in different applications. The cost of the technique would be divided by the number of applications.

Statisticians could offer the following to remote sensing:

(1) The support of the statistical information system (SIS) itself. It provides a framework for choosing areas of work.

(2) The SIS provides a dissemination structure allowing the target audience to be reached by means of all modern methods of communication.

(3) The SIS has a set of interlocking nomenclatures which have been internationally harmonized and are constantly being adapted. This experience of nomenclature work should be applied to remote sensing data so that they can be integrated in the best possible way into the rest of the statistical data.

(4) Hybrid applications will be developed. The cost of remote sensing will be lower when it can be applied where the comparative advantages are clearest.

(5) Integration in the SIS will allow remote sensing to develop multi-purpose applications. In this way costs can be shared between the various uses.

(6) It will be possible to use the information already available in the statistical system in remote sensing applications (in particular for ground observation, which will be combined with the results of traditional surveying).

(7) Statisticians have acquired special skills which can be invaluable in the definition and implementation of remote sensing applications. For example:

- considerable experience of computerization in developing and installing technologically complex applications;
- great skill in presenting results in the form of texts, tables, graphs and maps with summaries adapted to a wide variety of audiences;
- statisticians are being increasingly called on to develop the geographical dimension of their work. This makes them ever better able to situate the results of the remote-sensed applications in a general spatial context.⁵

A STRUCTURED APPROACH

'The arrival of the first usable results of the MARS project is enabling us to make rapid progress.

Mastering the tool

What measures should Eurostat take to develop the specific qualities of these promising tools?

A Working Party on Crops has been set up in conjunction with the Member States and, in addition, we are monitoring the implementation of action 1 ("regional inventories") of the MARS project.

These inventories will enable us to compare experiences and pinpoint any danger of non-convergence, by using both traditional and new observation methods.

In addition, in the interpretation of the image there is still great dependence on experts whose specialized knowledge in their sector is essential for proper interpretation of the earth. The development of expert systems should be able to compensate for this dependence.

The "spatial statistics" task force

The "remote sensing" section of this team has the following tasks:

- to monitor the work in progress in the area of remote sensing;
- to prepare a plan allocating responsibilities for the monitoring and management of specific actions;
- to examine the links between data resulting from remote sensing and the geographical information systems;
- to propose structures and procedures that will enable Eurostat to monitor the general impact of remote sensing on statistics and to take any necessary action;
- to organize the seminar with statisticians from the Member States involved in remote sensing.

Other developments

The other activities followed closely include:

- the work of the space policy unit of DG XII, a large part of which concerns the observation of the earth; we are much involved in the work of this Directorate, which is now producing very concrete results;
- the work of the joint Commission-European Space Agency Working Group on the Environment and Earth Observation.²

PREPARING FOR THE FUTURE

A seminar is soon to be held on all our work in this area. The subject will be: the impact of remote sensing on the European statistical system.

The date: 22-24 September 1992

Organized by Eurostat, the Netherlands Central Bureau of Statistics (CBS, Voorburg) and the Italian National Statistical Institute (ISTAT, Rome) in conjunction with the Institute for Remote Sensing Applications, Ispra, and DG XII, Brussels, the seminar will discuss how official statistical organizations can and should react to the new possibilities opened up by remote sensing.

The seminar will be aimed at both Community and EFTA countries. North America will also make a contribution.

The venue and other organizational details will be finalized shortly.

¹ First announcement of the seminar to be held from 22 to 24 September 1992.

² Interview with Mr D. Heath on 7 February 1992.

³ *Echos de l'Europe*, No 1-1992, p. 5.

⁴ Conference on 'The application of remote sensing to agricultural statistics', Varese, 10 and 11 October 1989 – introduction by Mr D. Heath.

⁵ 'Application of remote sensing to agricultural statistics', Belgirate, 26 and 27 November 1991 – introduction by Mr D. Heath.

S

TATISTICS AT THE DAWN OF THE INFORMATION AGE

Interview with Mr Roger Cubitt

The 21st century will be the age of information. For Eurostat, this will mean moving away from the idea of 'statistical data' towards that of 'statistical information'. The comments of an information technology specialist, whose background is neither that of an economist nor that of a statistician, highlight some of the trends likely to have an impact on future developments.

After graduating in computer science in 1971, Roger Cubitt worked in a variety of areas, such as naval architecture, transportation planning and chemical pathology, before being recruited by Eurostat in 1979.

At the Statistical Office, he was initially responsible for the development of software for the Cronos database system and took part in a feasibility study for Eurostat's future statistical architecture. Prior to becoming Head of the 'Database management and publications' Unit, he was responsible for coordinating Eurostat's major database development projects.

In 1986, he was the general chairman of the Third International Workshop on 'Statistical and scientific database management' in Luxembourg.

FROM DATA TO INFORMATION

Users of statistical information are no longer satisfied simply with figures: they want to know what they represent. This is happening in all areas where demand is growing – at exponential rates!¹

Reasons for the growth

Data series are becoming longer: over 600 000 new terms are entered in Cronos each month. This does not mean that the old data are removed, as that would entail losing the value they acquire as they cover longer and longer periods. In addition:

- 'in sectors already in existence, much more detailed information is required;
- statistical information is being called for in a number of new sectors, such as international negotiations (GATT), the revision of nomenclatures, the revision of many surveys at more frequent intervals, etc.'²

This explosion in demand could overwhelm those working in information technology. In such a situation, it is increasingly difficult to provide users with a clear idea of what information is available and how they can gain access to it.

THE PROBLEM OF DESCRIPTORS

The constant expansion in demand for data has led to an increasing need for descriptors. It also means that descriptors have to be as standardized as possible.

Descriptors are arrived at in two ways:

- The 'official' route:

A typical official standardization process is that of the nomenclature, which is created at European and world level. Developments in this area have so far

been extremely slow: transition into use can be a matter of decades;

- The 'pragmatic' route:

Here changes can be swifter. Dictated by the demand from organizations that need figures for their negotiations, this type of process can lead to rapid, sometimes disorganized definitions; some have to be fundamentally reviewed when an international definition is agreed.

'One thing is clear; the coverage in terms of agreed descriptors for statistical data ... is growing, as the complexity, detail and volume of available data expands.'²

Statistics for the man in the street

Legislative or quasi-legal documents increasingly refer to statistics. When GNP is used as a basis for wage negotiations, the users – who are not necessarily statistical experts – need to be told in simple language how the statistics presented to them have been produced and what they mean.

Geographical meta-data

This is one example. At the moment, statistical data are closely linked to official reference frames such as communes, boroughs, countries, regions, etc. The new geographical information systems (GIS) should enable this information to be placed in a 'space' perspective. 'This trend is already evident and is posing a number of problems as to how best to match the statistical information systems (SIS) with the GIS in functional and technical terms.'²

The systems of analysis

Techniques that were unusable 10 years ago can now be used on portable lap-top computers. There is no longer any need to reserve the data room for a week. But here, too, with all these packages for analysis available, further work

ought to have been carried out on data typology and some form of explanation should have been provided on the general outline of the information, so that the person about to tackle an application can use the right tool.

As regards GIS, again, new systems of analysis should enable better use to be made of geocoded data.

In the area of application-driven technology, the extent of the changes that the CD-ROM may produce on the market cannot yet be gauged. The arrival of products containing 500 megabytes of statistical data, together with an incredibly user-friendly software, all at an extremely accessible price, is bound to have a very significant impact on all areas of information, including statistical.

Confidential!

One final significant trend that should be mentioned here concerns confidentiality. As individual data are becoming more easily manipulable, so democratic countries are trying to protect such data more effectively. Why is confidentiality a problem?

- 'Statistical data requirements are becoming more detailed in both classification and geographic terms. This makes it more and more easy to identify individual respondents.
- The budgets for statistics are not keeping pace with the demand for statistical information; data are sometimes sought elsewhere, in conditions which do not comply with those that official collecting agencies are bound to observe.
- Pre-aggregated data rapidly become outdated, given that demand is beginning to change considerably; this leads to the breakdown of aggregated products into several sub-products, from which individual data may "resurface".²

THE DEVELOPMENT OF THE SIS

What do all these trends imply?

At the functional level

In the medium and long term

The public wants information, not figures. Radical changes are needed in the over-abstract modelling mechanisms. 'Complex oper-

ators' are information operators and no longer simple data operators. They will be based on object-oriented systems. These are computer systems which attempt to model objects and give them attributes and a set of functions in terms of properties; the aim is to provoke 'reactions' between well-defined entities.

For instance, if the concept of GDP were placed in conjunction with another phenomenon, such as unemployment, this would provoke a 'reaction' which the system would be capable of describing in a way which, ideally, anyone should be able to understand.

Experience has so far shown that statisticians should create their technical instruments themselves. Statistics is only a small part of the market for hard- and software producers. 'We therefore have to take the tools that exist on the market, transform them and combine them with the systems that we have developed ourselves. The development of software and the improvement of models will enable us to offer the user much more powerful and more transparent products.'¹

In the short term

We are condemned, in the short term, to continuing piecing together a variety of tools to meet SIS needs. Paradoxically, the user has the impression that everything is becoming simpler, whereas many things are in fact becoming more complicated. The proliferation of PCs, the harmonization and availability of teletransmission, the possibility of working from menus, of selecting services, of working simultaneously on different computers and communicating between them, do not give any idea of the vast complexity of what is going on behind the screen!

Let us take an example

Eurostat disseminates the main macroeconomic indicators of the major statistical sectors via the Eurocron base. Given the high level of use made of this, it has been decided that added to this general base there should be a more comprehensive base containing all the 'primary' information used to compile the succinct summary tables presented in Eurocron. This new base is called Infocart.

Objectives: Infocart is to function in two ways:

- When consulting Eurocron (Videotex), the user sees short time series (presented in a succinct table covering, for instance, a few months or quarters). He will be able, at any moment, to switch over automatically to Infocart, which will enable him to consult the full series (all the periods available) and also provide access to a greater wealth of documentation describing the data (explanatory notes, definitions, etc.). He will also have tools to 'download' the data to his PC so that he can include them in his own work environment.
- It will also be possible to use Infocart as a database for independent dissemination (which can be consulted outside the Eurocron context), offering search and downloading facilities.

The technology behind this project is highly complex (using one PC, parallel consultation of two completely different databases on different machines!).

Infocart is the first Commission dissemination base to use 'client-servers' technology. This involves the following:

- data are stored in a base installed on a central computer (Oracle);
- the user interface (which enables the user to carry out searches and consult the results, known in jargon as UFI – user-friendly interface) is a software installed on a PC;
- communication between the two machines is carried out in the following way: when the UFI requires information, it transmits its request to a software (also installed on a PC) which, in telecommunications, plays the role of 'client'. The client interprets the request, formulates it and sends it to a software called the 'server';
- the 'server' (installed 'somewhere' on the network, not necessarily on the same equipment as the Oracle base) receives the request, analyses it, 'discovers' that it must consult the Oracle base in order to reply, creates consultation transactions with Oracle, awaits the reply, receives the reply and redirects it to the client;
- the client acknowledges receipt and delivers it to the UFI.

Why so complicated? Because, in this way, using one PC (the UFI) it will be possible to consult x number of Oracle-type bases (which may be on the same machine or on different machines) in an entirely 'transparent' manner for the user, giving the impression that he is using just one base!

In view of all these aspects, Infofact is an experimental project which has been used to create a telecommunications infrastructure.

Make the machines talk!

Can short-term solutions be expected in the field of meta-data? The main risk is that statisticians will not manage to keep pace with events and set up the classification typologies needed to satisfy the demand for sending messages electronically, with the required textual elements. Commercial interests are developing more rapidly than institutional involvement in this area!

Work on Edifact³ will gradually help the sender and the receiver to understand and correctly interpret the content of messages.

Strategic aspects

Networks have to be integrated into the SIS architecture. In some cases, remote processing could provide a solution to certain problems of confidentiality.

Moreover, statistical and economic information increasingly has to be looked at in a 'whole world' context. It is clearly impossible to store the whole world's information in one place! This also applies to processing. It will probably have to be based on a combination of regional and national structures. High-speed networking technologies and packet communication systems will continue to promote decentralization of the analytical work of the economist, statistician and research worker towards their workstations, with the source bases continuing to require constant updating. However, the user interface at the workstation will, of course, have to be syntactically uniform, irrespective of the data source and processing environment being used.

MAINFRAME COMPUTERS TO BE REVIEWED

The mismatch

Data are collected by sector, the sector definitions being thematically targeted. Products may be sectoral but a large number of them require input from a wide variety of statistical sectors. How can collection and dissemination be properly coordinated if those working in the former lack knowledge of the product that is to be created and those in the latter lack knowledge of the data available?

Apart from this, the security and harmonization aspects will be even more difficult to implement in the 'distributed' environment.

'One – partial – possible response would be a "warehouse" type model. Information entering is already "pallet" size; these pallets are stored intelligently, without being disassembled; when information is requested, it is delivered with the whole of the pallet(s) on which it was originally packed, delivered and stored. It would be up to the user to "unpack" it and find the sections that interest him. It would be our job to ensure that the pallets delivered contain (at least) everything that is requested.¹

The content of the pallets would have to be properly labelled with descriptors. And what about security? And reliability?

'The picture is one of a decentralization of the SIS functions linked to the evolution of basic nomenclatures and meta-data availability. The time-frame, however, is to be measured in decades!²

A NEW ARCHITECTURE

Eurostat is radically reviewing the architecture of its information systems in order to create a fully-fledged European statistical system.

The main trend seems to be towards the separation of production systems from user-orientated and dissemination systems.

It appears that a buffer zone, known as the 'reference' layer, is needed between production and dissemination. The purpose is not simply to better ensure functional and data independence but also, and more importantly, to provide

some sort of global directory, or catalogue, which describes the data Eurostat has at its disposal and in which product(s) they are used.

The construction of this architecture and the incorporation of this reference layer will undergo considerable development and is a subject to which we will certainly return.

¹ Interview with Mr R. Cubitt on 31 January 1992.

² 'Statistical information systems into the 21st century' by R. Cubitt (paper for the sixth Conference on 'Scientific and standardized database management', 9 to 11 June 1992, Monte Verita, Switzerland.

³ See the article on page 20.

M ISSIONS, VISITS AND MEETINGS

■ Statistical cooperation with the Baltic States

At the beginning of January a meeting was held in Helsinki with the participation of the Directors-General of the national statistical offices in the Member States, and representatives from CSO/Finland, Statistics Sweden, EFTA and Eurostat.

A first evaluation of the statistical situation and requirements in the three Baltic States had been made during a mission of representatives of the EFTA countries and the EFTA Secretariat in November 1991.

The meetings were held in accordance with the Eurostat–EFTA agreement, under which EFTA is to prepare the ground for a joint EC–EFTA statistical cooperation programme for the Baltic States. A very comprehensive report on the situation in these countries had been drawn up by Mr Snorrason. The aim of the January meeting was to prepare the EC–EFTA assistance programme.

Draft cooperation programme

This first draft covers the following domains:

- (a) National accounts;
- (b) Price statistics;
- (c) Foreign trade statistics;
- (d) Agricultural statistics;
- (e) Financial statistics;

- (f) Labour force statistics;
- (g) Business registers;
- (h) Basic statistics;
- (i) Statistical data-processing;
- (j) General assistance.

Each of projects (a) to (g) consists of two sets of actions: basic theoretical training and on-the-job training.

The Baltic States are to ask the IMF to coordinate project (e) within the framework of the procedures laid down for their membership. Should the IMF not accept this role, coordination will be assured by EFTA.

Project (h) will deal with general and specific training (statistical standards, sampling techniques, etc.).

Project (i) will cover all aspects of the data-processing plan.

Project (j) will comprise two main sections:

- a basic seminar, to be held on a recurrent basis, concerning the role and operation of statistics in a market economy, and supplying the conceptual, institutional and economic background for the other projects;
- on-the-spot continuous assistance by non-specialist experts (project coordination, solution of day-to-day problems, administrative back-up, etc.).

■ Approval of the Comedi project

At the end of January 1992 the Commission approved the measures to finance the Comedi project. Launched in June 1990 to support and facilitate the implementation of Intrastat, the aim of this project is to establish an infrastructure which will simplify the collection and processing of statistics on intra-Community trade.

The chief measures provided for in the Comedi project are as follows:

- to run an information, training and awareness campaign among businesses;
- to help the national systems to adapt;
- to carry out adaptations within Eurostat (in particular the Comext-93 project);
- to carry out feasibility studies on a collection infrastructure;
- to conduct a series of trials in the field of EDI and new technologies.

T HE MANY CHALLENGES OF EDI

Interview with Mr Philippe Lebaube

'Developments in information technology and especially telecommunications have led to the introduction of EDI (electronic data interchange), which will enable the information systems of businesses and government departments to communicate during the main stages of commercial and administrative activity (e.g. production, distribution, invoicing, customs declarations, etc.). The gradual introduction of EDI and the advent of the paperless society will make the task of statisticians increasingly difficult.'¹

An IT engineer graduating from the University of Grenoble, Philippe Lebaube began his career in a research centre in Grenoble developing software for use in the social sciences.

He joined the Commission in 1986 and is now a member of Eurostat's computing and statistics unit in charge of general services development (office automation back-up, logistical services for computing activities) and the introduction of new computing methods for statistics.

In this role, strongly oriented towards the outside world, he is chairman of the 'Message Development Group' of the Western European Edifact Board (part of the UN/Edifact organization). This group is responsible for ensuring that statistical facilities are as far as possible incorporated into EDI applications.

'The development of the modern economy forces statisticians to review collection processes for statistical information and attempt to integrate them into the very workings of businesses and government departments.

This challenge concerns statisticians working for national statistical organizations, but also business statisticians and economists and decision-makers in general, for whom the availability of effective statistical instruments is a *sine qua non* of management and planning.

EDI technology is developing rapidly in the more dynamic sectors of the economy: trade, transport, the motor industry, banking, chemicals, insurance, information technology, etc. Administrations are increasingly interested in EDI, both in their relations with businesses (e.g. the customs service) and for exchanging information with other government departments.¹

STATISTICIANS TALK TO LAYMEN

The need to communicate

Statisticians must break out of the ivory tower in which they only communicate with one another. They must get involved in the nitty-gritty of the wider world in which businesses and government departments communicate with their partners, with professional organizations, with statistical organizations or with collection centres, and this communication has to be a two-way business.

In the field of EDI, before 'communicating', one has to 'determine with one's partners the content of the information to be exchanged, and also, since it is the machines that are doing the talking, to determine the pattern of messages required to handle the transactions to be carried out.'¹

The right strategy

Statisticians have no authority to ask businesses and government

departments that have invested in Edifact to use another standard for exchanging statistical information. Hence our pragmatic approach: why shouldn't Edifact handle exchanges of statistical information in addition to its other functions?

The involvement of statisticians

'The involvement of statisticians in the design work for EDI systems is essential for the following reasons:

- the information systems of businesses and government departments are a priceless source of statistical information;
- the adaptation of information systems to EDI technology will imply a great deal of work on methods and models for data to be exchanged and, as a result, for data stored in the information systems of businesses and government departments.

These observations have two implications for statisticians:

- they must participate in the process of introducing EDI from the very start (modelling stage);
- they must use their expertise and experience in the field of information processing methods to enhance the contribution of statistics to EDI.¹

A promising standard

In the wider world described above, the approach adopted was to pick the standard on which there was already a degree of international consensus among users, in both the private and public sectors, for exchanging data structured with the aid of Edifact.'

Existing coordination structures

'Edifact is being organized by the United Nations through existing coordination channels. We are not starting from scratch: we are using existing structures, getting them to do their job, using the influence of these organizations to establish points of consensus that can be further developed in this type of forum.'⁴

Interchange agreements must be standardized

'The growing use of EDI has highlighted the need to standardize interchange agreements internationally.

The role of an EDI standard would be:

- to define the message structure to be used to organize and represent information;
- to list and document all messages that are harmonized and correspond to the requirements of users;
- to provide users with details of the ways in which the information is encoded.¹

The problems of flexibility

Though quite adequate in the wider environment in which users exchange high volumes of simple operational data, the Edifact standard is much less suitable as such for the statistical community. Statisticians have to exchange information that rarely has the same structure and their messages are therefore not as standard as those of businesses and government departments. They need to use very flexible standards that can handle time series or multidimensional matrices. SGML (standard general mark-up language), a mark-up standard much used in publishing, computer typesetting, etc., would have been much more flexible for representing variable structures, but it is not particularly suitable for numerical data, and little known among our partners.

Keeping the remit unchanged

'EDI technology must not affect the organizational remit of our partners. The system must be used by the organizations to collect information they are responsible for collecting. We must ensure at all costs that the new technology does not disrupt current organizational arrangements.'⁴

THE IMPACT ON DATA COLLECTION

How will EDI affect the collection of statistical data (bearing in mind that one man's collection is another man's dissemination)?

A paradox

While the demand for statistical information is growing rapidly, those responsible for providing the raw

data are increasingly reluctant to assume the 'burden of data collection', a task which becomes all the more onerous the better organized and better equipped they are with computer and telecommunications systems.

Faster and better

Eurostat is making every effort, through its initiatives on the Edifact board, to bring the whole statistical apparatus to an awareness of the need to work faster and in more detail in order to adapt to the modern economy's demand for fresher and more reliable statistics closer to the event.

The 'extended enterprise'

The modern economy constantly forces firms to be more competitive in the context of the 'extended enterprise' (namely such concepts as no-stock and just-in-time manufacturing, the paperless office, etc.) Such firms have management and organizational priorities to which statistics must necessarily adapt. We must not allow a technology gap to develop.

In order to be more reliable, statistics must get as close as possible to the operational level. It must attempt, for example, to draw its information directly from the accounting apparatus, etc.

Collection mechanisms

There is obviously no question of intercepting, like spies, messages that do not concern the statistician; but of installing collectors capable of drawing off, with due caution, information of use to the statistician while guaranteeing the right of confidentiality and secrecy.

The transitive model

Thousands of firms have a duty to provide statistical information, which they send to collection centres – professional associations, chambers of commerce, banks, central banks, customs organizations, NSIs, etc. All these flows of information must be identified and ways must be found of making them supply the statistical system as a whole.

'The general layout of these systems for collecting raw statistical data is such that the business or government department is rarely in direct contact with a statistical agency, which is the reason for the concept of a transitive model.'¹

The technology and organization gap

The possible existence of a technology gap raises the spectre of an organization gap. Even with a highly sophisticated EDI network, in technological terms, what if the organizations at either end of the wire do not agree on the concepts, the definition of the message, or on the frequency of exchanges?

'Things are much simpler in the operational world: objects, material quantities, are easy to define. Statistics, which does not have the same operational requirements, must pass from an operational frame of reference to a statistical one, which is based on classification: in statistics the object is to classify, not to identify. It is at this point of transition that problems of definition and logic arise. Well-coordinated classification methods are therefore required – hence the role of statistical standardization – to build the necessary bridges between two worlds with very different priorities.'⁴

Partnerships must be well defined

This will be the focus of a whole programme of coordination work. Without infringing the organizations' remits, exchange agreements will have to be reached that are beneficial to both parties. 'It is difficult to establish a partnership by saying: "give me your confidential information; we promise not to disclose it outside the statistical context for which it is required". Far better to say: "keep your confidential information; when I need access to that level of detail in order to gross up the data at European level, for which we need to produce aggregate figures, I will go and get it myself, and you will obviously have the benefit of the aggregate figures that your information has helped produce".'⁴

In short: the position of statistics in EDI systems is the transitive model

'The reasons why statisticians should get involved in the development of EDI are the following:

- the information systems of businesses and government departments are increasingly computerized;
- these information systems are potential sources of statistical data;

- businesses and government departments are getting involved in EDI because it facilitates their transactions with their various partners;
- for statistics, the methods for collecting data directly from businesses and government departments ought to be based on the same technology they use in their operations;
- businesses and government departments will be more willing to provide basic information if they receive statistical information in return.¹

From single-sector to multi-sector systems via standard protocols.

The first users of EDI were in the same sector: bankers talking to bankers, motor manufacturers or chemical companies talking to other motor manufacturers or chemical companies, etc. Today EDI is becoming multi-sectoral and statistics must adapt to all contexts in which it must make observations, but without having to define a different approach or a different type of message for each sector.

'Depending on the type of activity the various users are engaged in, current developments in EDI aim to involve partners from different sectors in EDI exchanges.

Businesses wish to communicate via EDI with their subcontractors and suppliers, but also with their bankers and hauliers, and with their supervisory government departments.

This multi-sectoral trend is only possible if the EDI technology is based on standard protocols.¹

THE IMPACT ON DISSEMINATION

A universal standard does not exist

In theory, dissemination is the mirror image of collection. EDI technology used to collect data must also be capable of being used to disseminate data. 'But we have to be multimedia: we need to produce not only statistics, but graphs and comments too, notes on statistical method and brief analyses, etc. And the Edifact standard is not suitable for that kind of thing at all.'⁴

Multimedia

'That is why we are interested in new technology suited to this kind of exchange that is multimedia both in terms of the data-media used and in terms of content. In its role as a transmitter and receiver, Eurostat will in future be in a position to collect as well as disseminate multimedia products involving text, tables, graphs and illustrations, as well as numerical data.'⁴

The Strings project

The aim of this project is to create a multimedia environment for the production and dissemination of statistics. The set-up should ensure that from the same observation we obtain the same result whatever data medium it was received on: hard copy, Videotex, CD-ROM, SGML electronic file transmitted from outside, etc.

'As part of the Caddia programme, Strings has already enabled us to carry out interface standardization trials between the applications environment and the electronic publishing environment, including communications with the exterior. Eurostat's aim is to provide common infrastructures and standardized representation for multi-format data compatible with all the media used to disseminate statistics.'⁴

BREAKING DOWN THE BARRIERS

Statistical systems must get as close as possible to the source of the information, without interrupting normal procedures, in order to speed up the intermediate processing stages. Even the manner in which information is stored has to be much closer to the remit given to the collecting organizations.

A different concept of the future

Exchange channels must be set up between collection and dissemination. The technological challenge is to design applications that enable the various participants to work together in partnership. But we have to discover the preconditions, in terms of technology, organization and standardization, for breaking down the barriers surrounding statistical systems.

Leave responsibilities where they lie

In the new statistical applications, the systems will have to go and fetch the information where and when it is needed, leaving the management responsibility where it lies, with those who have the remit, making the best use of each person's role at his own level.

Distributed statistical information

As an extension of the work of the Telematics and Statistics Task Force, set up following the November 1990 DGINS conference, the concept of a distributed statistical system is being developed.

DSIS (distributed statistical information system): a conceptual framework

'The national statistical institutes (NSI) are different in several respects: their size, their information systems, their organization (structural and operational) and their control of the collection of information. They have their own traditions, organizational methods and remits, which should be respected. This means that they will only participate in a system if it respects their independence.

A European system of statistical information will be based on:

- the telematic interconnection of the NSIs' information processing systems;
- applications-level interoperation standards;
- well-structured concepts to promote common services;
- the pooling of results and the non-duplication of effort;
- common technological progress in order to offer competitive services in an open European information market.

In order to develop the dissemination of various types of information and services, a conceptual framework needs to be devised to provide common objectives and standards, and to harmonize concepts and methods. The development of this conceptual framework will be closely linked to the harmonization work done on statistics at European and international level and will be related to cooperation agreements established between statistical organizations.'³

What we are attempting to set up is not a monolithic system; on the

contrary, it is a context, a framework. It means being able to say: 'for such and such a collection requirement, use such and such a technology, establish such and such a protocol, etc.' A major study is currently being carried out in the Member States to define the technical and organizational preconditions for this type of architecture to work.

THE WESTERN EUROPEAN STATISTICS GROUP: A EUROSTAT INITIATIVE

In November 1989, at Eurostat's initiative, a statistics group was set up within the Western European regional board of the UN/Edifact organization. The main objectives of this group are the following:

- to take charge of the standardization process for statistical messages;
- to study the introduction of collection systems for statistical information drawn from other standardized messages;
- to define standardized messages for exchanging statistical information.¹

To carry out this work, the group has set up five subgroups (WGs – working groups), each under the authority of a coordinator. Relations between the work of the various groups are dealt with at plenary meetings of the group under the chairmanship of Eurostat.

WG1: Exchanges of aggregated statistical data

'This group concentrates on the definition of messages for the exchange of aggregated statistical data such as time series and multi-dimensional tables. Such messages must contain not only the basic data (codes, headings, figures) but also related information – meta-information – for the interpretation of the data (origin, statistical methods, dimensions, notes, etc.). These messages are intended to be used in collection and dissemination operations between grass-roots reporting units, statistical bodies or preferential users (e.g. disseminators of statistical data).

WG3: Statistical aspects of code lists

The aims of this group are:

- to coordinate and harmonize coding systems for statistical information in EDI messages, especially through the use of statistical classifications;
- to encourage the use of statistical classifications in other EDI messages;
- to define the pattern of EDI messages for the dissemination of statistical classifications between agencies responsible for updating (e.g. NSIs, Eurostat, OECD, UN/SO, users, etc.);
- to carry out standardization work to introduce statistical classifications into standard directories (See Untdid: United Nations Trade Data Interchange Directory).

The group will pay special attention to the multilingual treatment of headings, annotations and notes on statistical methods related to the classifications.

WG4: New ways of collecting raw statistical information

The principal role of this exploratory group is to identify statistical fields and projects for which EDI technology is particularly appropriate.

WG5: External trade statistics

This group studies procedures for collecting and disseminating statistical data on external trade. The aim is to enable collection centres to receive messages from grass-roots reporting units containing statistical information on intra-Community trade after 1992 and on trade with non-Community countries.

WG6: Balance of payments statistics

The objective of this group is to design messages for the compilation of balance-of-payments statistics. The partners in this project will be the grass-roots reporting units, banks, central banks and NSIs. The basic idea is to include the information necessary for drawing up the balance of payments in the financial messages used to carry out international payments transactions.¹

TOWARDS A GREATER DEGREE OF INTEGRATION

The world of operational EDI is gradually becoming more integrated. Will this extend to the exchange of multimedia statistical information?

Different perspectives on the same objectives

'This strategy will draw on other exchange standards, such as SGML, mentioned above, or CALS (computer-aided acquisition and logistics support – an initiative of the US Department of Defense – DoD). Transfers will involve:

- data,
- documents,
- technical notes,
- invoices, order forms, invitations to tender, etc.⁴

Intersecting frameworks

For information systems handling these different types of data, the frameworks necessarily intersect. It is difficult to imagine working with different conventions to identify the same operations. We have to identify points of convergence and hammer out a uniform framework. All the organizations concerned must realize the vital importance of incorporating into their information system suitable interfaces enabling them to coordinate their communications with the outside world.

A more coordinated technological approach

There is no point in distributing and devolving operations without a well-coordinated data framework: the edifice would become a 'tower of Babel'. 'We must not get carried away with technology for the sake of technology, we must understand its effects on organization. What is happening today with EDI must enhance the awareness of national government departments to the importance of concerting their technological approaches.'⁴

CONCLUSION

'Experience gained with the "EDI and statistics" project and the work of the WE/EB-MD6 show that statistical standardization and EDI standardization go hand in hand. The latter must not go be-

fore the former, but must be its language, its vector. What is more, EDI standardization can throw light on the difficulties of establishing a framework for statistical standardization

Projects please!

(As for action) ... it seems better to put the emphasis on a pragmatic, project-based approach. Participation in these projects will vary depending on who is involved and what the regulatory basis and geographical context are. Broadening the participation in concrete projects will be easier to attain than attempting to establish a universal standard from the outset.¹

A word to the wise ...

¹ EDI and statistics – The challenge for statisticians – Ph. Lebaube, December 1991, Eurostat/A1/NT91-45

² UN/Edifact ISO-9735 – United Nations Rules for electronic data interchange for administration, commerce and transport

³ DSIS (distributed statistical information system framework) – State of understanding, March 1992, Eurostat/A1/NT92-12

⁴ Interview with Mr Philippe Lebaube, 31 January 1992

NEW TECHNIQUES AND TECHNOLOGIES FOR STATISTICS

Interviews with Daniel Defays and Photis Nanopoulos

The aim of the seminar organized in Bonn (24 to 26 February 1992) was to explore the opportunities for statistics created by the new technologies. A further objective was to help identify the fields in which progress should be encouraged and thus contribute to the development of new research programmes for the years to come.

Daniel Defays

Holder of a Ph.D. in mathematics and statistics from the University of Liège, he has a special interest in psychology and the quantitative methods used in this field. A teacher and researcher at the Faculty of Psychology and Educational Science in Liège, where he still lectures, he joined the Commission in 1979. He spent a year in the United States in 1986 at an artificial intelligence research laboratory.

He currently heads Eurostat's 'Research, development and statistical methods' unit.

Photis Nanopoulos

A graduate of the Engineering School of the Greek Airforce, he studied mathematics and physics at the Strasbourg Institute of Mathematics, and computing and statistics at the University of Nancy. He travelled to the United States, where he gained a doctorate in statistics at Berkeley. On his return to France he studied for a doctorate (probability and number theory) at the University of Strasbourg, where he also taught. After teaching at the Athens Polytechnic he joined Eurostat in 1983 and is currently Director of Directorate D: Business Statistics.

NEW TECHNIQUES AND TECHNOLOGIES ...

Technology is the study of techniques. Including both terms in the title has silenced any objections as to terminology. 'But you still have to agree about what really are the new techniques in statistics today:

- the use of increasingly powerful computers, for example computers working in parallel;
- networking, the development of telecommunications facilities;
- the use of disciplines such as artificial intelligence, which provide scope for representing knowledge and not just information;
- new methods for disseminating information such as CD-ROM and passive or interactive videodisks. These new techniques are obviously going to transform the task of the statistician;
- all the new opportunities created by computers. The computer is no longer a mere substitute for tasks previously carried out by hand, today it changes the nature of these tasks. Techniques for scanning data bear no relation to the way things were done before. In analysing data, the statistician can use graphic techniques that situate points in two or three dimensions. He can rotate these groups of points, superimpose a histogram and a graph on cartesian coordinates ... a whole series of new opportunities that considerably broaden the range of instruments available. With the use of machines, we can now exploit man's extraordinary capacity to recognize shapes;
- all the techniques of remote sensing, which work with extremely precise spatial coordinates that are highly disaggregated geographically. While statisticians are familiar with working on time series, they now have to deal with data that have not only a temporal dimension but a spatial one too. A

number of systems are emerging which, by stratification, superimposition, etc., will enable statisticians to throw light on phenomena that could not be represented before, using data that are difficult to grasp intuitively.³

... AND THE BENEFITS FOR ENTERPRISES

Easing the burden of collection

'The techniques that are being developed will enable existing sources of information to be used to better effect without necessarily forcing businesses to restructure their information sources to suit the need of statisticians.

The manager will be able to describe his activities in non-structured terms, in everyday language: the new systems will be capable of receiving this type of message and transforming it into the kind of structured information the statistician requires.

Statistical surveys can also be accompanied by a whole range of services that did not exist before, with a user-friendliness and efficiency that will greatly simplify life for businesses.³

Facilitating access to information

In exchange for the information they supply, businesses want access to information on other businesses and other sectors. This 'return information' could become much more personalized, faster, more efficient and more systematic.

'Current efforts to develop the concept of meta-data will make it possible to invest numerical data with a whole series of comments and information that will help businesses to interpret the statistics and ensure that they are well employed.³

Helping them to master it

'With information becoming increasingly widespread thanks to EDI, and with all the facilities that will become available, businesses are in danger of being swamped with information. The problem will be to check the validity of the information and separate the wheat from the chaff. New statistical techniques could make it possible to carry out validity checks and obtain a clearer idea of where the relevant information is to be found.'³ 'Statistics involves a whole series of processing operations; the main thing is to know at what point in the chain businesses should intervene to obtain the information they require.'²

We must also act to ensure that large and small firms can compete on more equal terms. 'The development of expert systems, which enable the user to carry out increasingly complex and detailed analyses hitherto the preserve of large businesses, will enable their smaller counterparts to use the information too, and will permit all partners to compete on more equal terms.'²

SMEs too

More widespread access to information obviously means easier access, especially for SMEs. Much 'strategic' information has hitherto been confined to the major firms, which alone had the necessary means and resources. 'Obtaining the information is not enough. If the firm does not have teams capable of processing it, the information is largely wasted. Increasingly the information will have to be supplied in a "pre-processed" form.'²

THE ORIGIN OF THE SEMINAR

Incorporating statistics into research

'The Community's remit in the area of research and development (R&D) is very broad. Various programmes are in progress and, in 1987, we succeeded in having statistical activities included in one of these. This gave rise to the Doses programme⁴, which is aimed at preparing the ground for the statistical information system of the 21st century.

Identifying an initial technology

Faced with the dilemma of how to produce more information without overloading respondents (especially businesses), statisticians must modernize their instruments and methods. This modernization cannot always be carried out in a step-by-step fashion: innovations have to be introduced, and the ground has to be prepared for them.

As the Commission's research programmes are structured by technology, it was necessary to start with the instruments and not with the problems to be solved. For statistics it was therefore necessary to present a programme based on a particular technology. Our choice fell on a technique that appeared particularly promising at the time: artificial intelligence.'³

A good definition

'This is a technique that involves getting a machine to do things that are supposed to require intelligence.'³ 'It is a new way of dealing with information and organizing it: not only the data themselves, but knowledge about the data and about ways of processing them. The aim is to integrate more completely into our systems procedures for linking microeconomic data to macroeconomic data, to models of use, inference, forecasting, etc. and to incorporate techniques that can be used with great efficiency without danger of disclosing information covered by statistical confidentiality.'²

A vast field of application

So far in statistics, virtually everything to do with the digital processing of information has become almost fully automated. But the statistician is faced with a whole series of tasks that, while they are not digital, are perfectly systematic and conventional (coding activities, products, etc.). With the application of artificial intelligence techniques it will become possible to computerize all these exceedingly time-consuming tasks.

'The main expense of new technology lies in the increased training costs that it always entails. And as technology evolves very quickly, the investment in training is quickly lost.'²

A wide range of concerns

By including statistics in research programmes and attempting to encourage the use of expert systems, Doses aims at a wide range of objectives:

- preparing the instruments that will be needed in the next 5 to 10 years;
- involving researchers in types of problem that they are not familiar with;
- encouraging exchanges and co-operation between research teams from different countries;
- working towards cohesion: in some Member States statistical R&D investment is patently very limited. Helping such countries to work in more advanced teams contributes towards cohesion, even if it involves investing time which could usefully be spent on other tasks.

'As statistical know-how is very unequally distributed, anything that will help it to be transferred as quickly as possible to all the partners must be encouraged. But we cannot expect everyone to progress at the same speed ...'²

A REPORTING SEMINAR

'Several international teams have presented projects as part of Doses. The main aim of the seminar was to ask them to produce the initial conclusions of their work. And we wanted these conclusions to be presented to the scientific community and the NSIs.

Beyond Doses

We also wanted to go beyond Doses:

- by attempting to obtain contributions that had not been financed by the Community;
- by attempting also to prepare a future programme that would not be limited to artificial intelligence techniques, but would embrace all techniques that could be of use to statistics.'³

Three major themes

We therefore had to identify a number of technologies in which European investment would be of benefit to statistics as a whole. Our attention was directed towards three main themes:

- artificial intelligence;
- geographical information systems (GIS);

- telecommunications in general.

An excellently organized event

The scientific committees organized locally have worked wonderfully. From all corners of the globe, over 230 participants, mainly scientists, answered the invitation, which led to an extremely fruitful brain-storming session and to very useful informal contacts. GMD (Gesellschaft für Mathematik und Datenverarbeitung) proved to be a model of organization. The statistical offices of the Member States, as well as Statistics Canada, the US Bureau of Census and the ISI (International Statistical Institute) were fully involved organizing the event.

A place for Europe

Europe has a key role to play in this area for several reasons:

- 'the pace of integration within the European Community, or I should say the European Union, is accelerating. The demand for statistics at European level is growing exponentially;
- continued encouragements, stimulation and funding of research programmes in the area of statistics and technology are clearly needed: individual statistical agencies, research institutes, universities and software developers are unable or unwilling to invest enough in this kind of pre-competitive project, essential for progress and innovation;
- the exchange of expertise between the different countries has to be promoted in order to guarantee a better distribution of know-how and skill. The level of development of statistical systems in the various Member States varies widely. It is the role of Europe to stimulate international cooperation and to make expertise available to all Community users;
- coordination is also needed for reasons of efficiency. Establishing standards, for example, makes it possible to ensure the compatibility of future activities. Early consultations will allow everybody to benefit from existing experience, and to reduce overlaps and duplication of effort.¹

BUT REMEMBER ...

An enormous gap

There is a striking gulf between the world of research and that of public statistics:

- on the one side are researchers whose job is to produce systems, prototypes, models, to write articles, register patents, etc.;
- on the other are statisticians who want to be properly equipped to carry out their daily tasks.

One of Eurostat's roles is to bring the two sides together. 'Internationally all the advanced countries are working on it: a large number of individuals and institutions are forging new ideas which, as the seminar showed, have already led to the development of a number of extremely interesting products. Eurostat's initiative was enthusiastically received because the types of technique and technology referred to have hitherto always been developed by the innovators themselves, for whom – in our experience – the statistical point of view is the last consideration. We have made every effort to make them realize that they should pay greater attention to statistics, and I believe they have got the message.'²

More Value

The aim of the Community's Value programme is to identify the results of research programmes and bring them to market, to get software firms involved by saying: 'This is what researchers have come up with, would you be willing to develop it and produce a working system?'

'A different kind of mechanism needs to be established to bridge the gap between research and the world of official statistics, which is not really a particularly large segment of the market ... This is why the Commission has such an important role to play in this field.'³

Stimulating European research

'"What do you want us, a European public administration, to do to stimulate your research?" That is the question we asked the participants, but there were not many answers. The seminar was perhaps not the right place to ask it. We will get answers through other channels ...'³

BROADENING THE PROGRAMME

'The discussions enabled us to identify new directions for research. It should be noted that the approach varies depending on who you are speaking to:

- the NSIs have a problem-oriented way of looking at things: problems of data input, of data access for the user, of documentation, etc.;
- the researchers, however, come with solutions: one with neural networks, another with genetic logarithm, a third has been studying chaos theory ...

How do we bring together the problem setters and problem solvers?'

Different levels of research promotion

The Commission's methods of research promotion can be modulated:

- relatively low-key action: identifying ongoing research and supporting it financially, intervening in a fairly passive manner;
- getting a bit more involved: 'if there is a gap between the world of research and that of official statistics, let us engage researchers and send them on training placements to the NSIs so that they get a better idea of the problems of statistics and can apply their technical skills to them; in exchange, a number of officials can go to the NSIs to help them identify what action should be taken';³
- one step further: attempt to highlight areas of special interest, define the fields that concern statistics, and promote and finance research activities in these fields;
- further still: most NSIs lack the means and the resources to invest in R&D. 'In view of the need, let us pool our scanty resources and set up a statistical research centre.'³

'The Commission must promote discussion between managers and business economists on the one hand and statisticians on the other, in order to ensure that research leads to the creation of finished products that can actually be employed by the end user.'²

Let's not be too applications-oriented

Research that concentrates excessively on immediate applications quickly becomes sterile. The research must retain a disinterested character, even if this leads to a 70 to 80% failure rate: it is one of the conditions of creativity.

'Expecting users to determine the paths the research will follow on the basis of forecast needs is a dangerous business. In the days when people went around on horseback, if you asked what the "user" needed he would have replied: "I need better facilities for cleaning my stables". He would never have mentioned the car ... One can only express wishes on the basis of what one knows. If you had asked 20 years ago: "what do you expect from computers?", no one would have talked about interactive graphics ...'.³

Communication between researchers, who progress in a fairly spontaneous and chaotic manner, and between users, who identify their needs and attempt to take advantage of what is happening in the research field, is not an easy matter.

The main thing is to build bridges between the two worlds. This seminar has provided an opportunity for building more than one.

Coping with inequality

'In France there is the concept of the establishment or local unit. The statistician can determine easily that at a certain place such and such a quantity of such and such a product is produced or that so and so many jobs exist. In the Netherlands the concept is not the same: if you wish to know what is going on in a region, you have to contact all the businesses and ask them to tell you what they do. The two systems should logically reach the same result, but not at the same speed ... How are we to cope with these inequalities?

There is only one way: by developing equivalent instruments for producing statistics and improving some in preference to others.

There is no need to build an elaborate statistical structure in each country, but all countries must have at least the basic minimum for carrying out the work required at Community level by drawing in-

spiration from what is best in the inventory (which has still to be compiled) of instruments in use or being introduced.²

IN TERMS OF PROJECTS

After this account of what has been done so far, what does the immediate future hold?

Concerted initiatives

As regards stimulating artificial intelligence, there are still means of financing a few well-timed and concerted initiatives, provided they are European in scope and will in time be of benefit to official statistics. 'We have identified researchers who were doing the same research in different places round the globe. They are now going to concert their efforts. We also need to send young researchers to all the places where things are on the move. Could we incorporate this concern into the next statistical programme?'²

Intensifying the incorporation of statistics into EDI

'As the development of EDI systems gathers pace, not enough importance is given to statistics despite all our efforts. Additional resources must be allocated to this field. I am concerned about what transport professionals are going to do with transport statistics. The same applies to services. We are in danger of being caught out at a time when businesses are beginning to standardize messages covering exchanges of services. There is no time to lose.

If the standards now being developed do not take account of our input, changes will have to be made later, and this will cause great difficulties.²

ENS (European Nervous System)

What should be done in terms of computer networks to move towards a distributed statistical system in Europe without impinging on the independence of the Member States and yet ensuring sufficient central coordination?

Neural networks

'The origins are obviously biological. The aim is to produce a data transfer system based on the nodal points of a network, to provide the statistician with instruments that free him from the con-

straints of mathematical formalism, to get away from two-dimensional projections, to visualize things in a completely different manner with new models. It is a very promising field.'³

Geographical information and remote sensing

Beyond what has already been said, what new ideas are there in the area of quality control?

'Statistical information is often the result of an extremely long process. It has become strategic (e.g. GNP data, which are used in calculating national contributions to the Community budget).

As for its quality and validity, a thousand and one errors are possible, e.g.:

- systematic errors of coverage of the register;
- coding errors;
- omissions, or things one does not wish to disclose;
- random errors in sampling;
- other errors.

All these sources of error of various kinds combine. What is their effect on the final result? There is indeed a theory of statistical error – sampling error – but there is not yet a general error theory.³

'What now has to be done is to define projects that will enable statisticians not only to take advantage of new systems, but also to "perform" better, i.e. to produce statistics in a more readable and visually more attractive form.'²

Business panels

These would help in providing micro-data that do not constitute individual data and do not breach confidentiality, but enable the statistician to perceive possible changes in the economic landscape at an early stage.

'We would like to set up Community panels of between 50 and 60 000 businesses, both SMEs and large firms, chosen on the basis of their representativeness. Such panels could be used to produce data that would enable us to understand market trends as quickly as possible, a forecasting tool so to speak.'²

Registers

There have been many attempts – not all necessarily emanating from

the Statistical Office – to create a network of registers so that, for example, a business in one country that wishes to set up in another country can have easy access to market information.

Some countries are more 'forthcoming' than others. 'If we do not take the matter in hand, countries whose information systems are more open will tend to close up.'³

Accounting charts

'Linking the accounting chart with statistical concerns is another extremely important matter. We are in favour, but it is a very controversial subject ...'³

'Some would like to legislate, to impose an accounting chart. In my opinion this would be contrary to the principle of subsidiarity. Why not let standardization operate on its own? Company accounting software will increasingly bring about an interpenetration of concepts, occasionally contradictory, or will at least enable us to interpret data in the same way, whatever their source. What we must do is to get manufacturers to realize that statistical standards have to be incorporated into such programs and that it is their job to do it.

In the meantime, together with our counterparts in the various NSIs and in close contact with the work of DG XV on accounting legislation, we are attempting to ensure that when directives are translated into national law, statistical aspects will be included. In questionnaires, statistical data can be based on accounting concepts.

We are also planning, following the BAC concept (DG II's data base), to establish an individual database at company level linked to a balance sheet centre that would enable us to obtain more detailed data on businesses. Balance sheets are not sufficient to analyse the competitiveness of a business or a sector. We need to go further, to understand better the cost structure of the business at each level, the efficiency with which each factor contributes to the product, etc. This project will enable us to produce more accurate analyses of competitiveness.'²

A nomenclatures server

'The Member States are going to have to use the European classifications. When these are managed at European level and used at national level, a communications problem arises. It must be solved rapidly. This is also a unique occasion to link this service with a whole series of other services. The new technologies are forcing us to reconsider our role. Once we are using the classifications, a whole series of new tasks that did not exist before will have to be tackled.'³

OK ARISTOTLE

'Statistical techniques and technologies are unequally distributed in the countries of the European Union. The "search for research" carries the statistician to the four corners of the globe. There must be a gigantic transfer of know-how – and quickly – overcoming the barriers of language, culture, etc., and scientific and computing standardization is what will make it possible.'²

'Europe cannot be strong without mastering information technology from an industrial and technological point of view. There is no reason why Europe, which gave birth to humanism – the alliance of science and consciousness through writing and the printing press – should not be the creator of a new humanism grounded on the use of information and knowledge. We have the expertise, what we need is the will, the determination and the courage to see it through. As Aristotle said: "The more difficult a thing is, the more it calls for art and virtue".'⁵

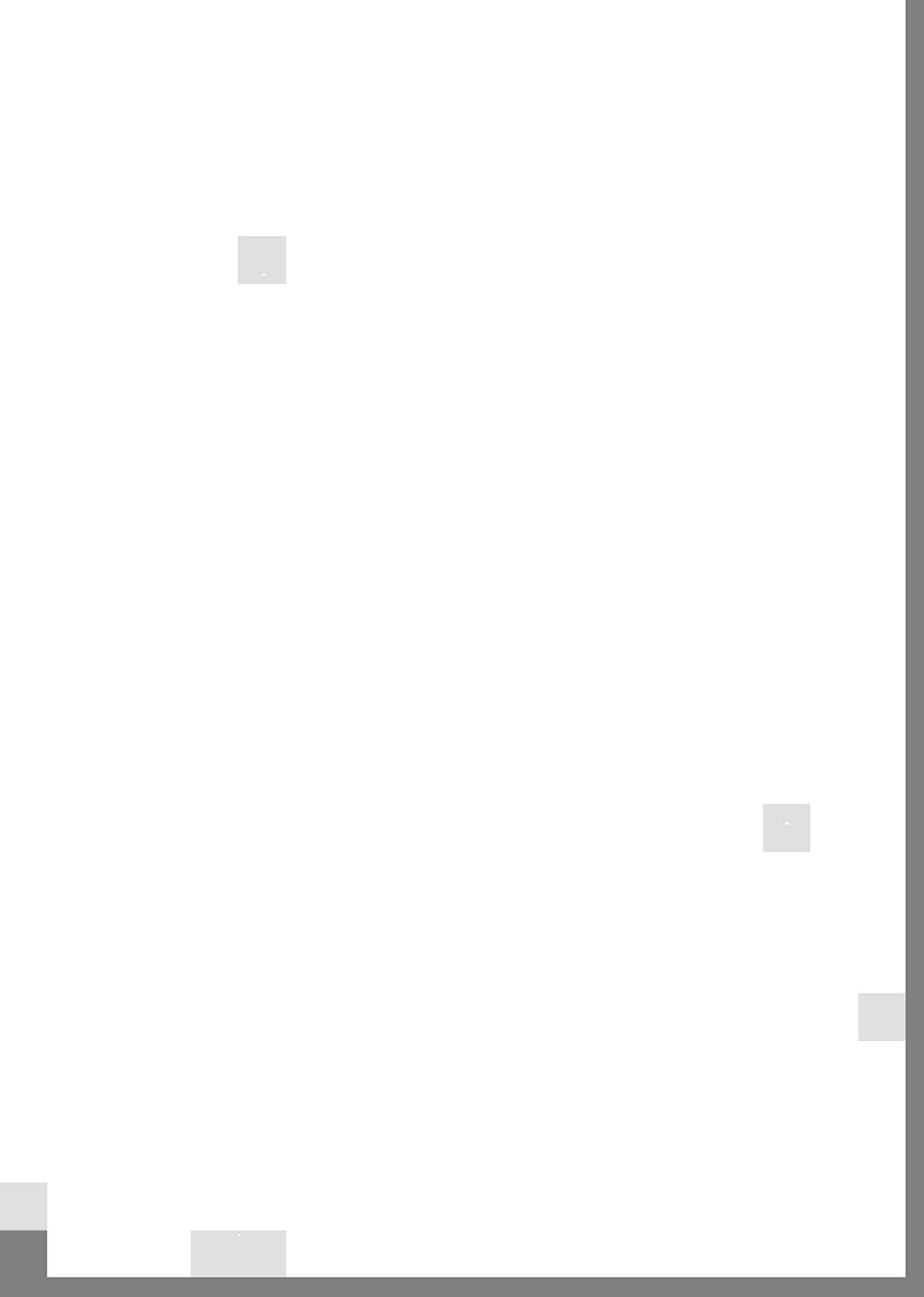
¹ Opening speech of the Bonn seminar given by Mr Yves Franchet on 24 February 1992.

² Interview with Mr P. Nanopoulos, 27 February 1992.

³ Interview with Mr D. Defays, 27 February 1992.

⁴ Doses: Development of statistical expert systems programme.

⁵ 'Les enjeux de l'informatique européenne' by Jacques Stern, Chairman of the Bull Group, in *L'entreprise et l'homme*, No 4/88, p. 145.



Venta y suscripciones • Salg og abonnement • Verkauf und Abonnement • Πωλήσεις και συνδρομές
 Sales and subscriptions • Vente et abonnements • Vendita e abbonamenti
 Verkoop en abonnementen • Venda e assinaturas

BELGIQUE / BELGIË

Moniteur belge / Belgisch Staatsblad

Rue de Louvain 42 / Leuvenseweg 42
 B-1000 Bruxelles / B-1000 Brussel
 Tél. (02) 512 00 26
 Fax (02) 511 01 84

Autres distributeurs / Overige verkooppunten

Librairie européenne/ Europese boekhandel

Rue de la Loi 244 / Wetstraat 244
 B-1040 Bruxelles / B-1040 Brussel
 Tél. (02) 231 04 35
 Fax (02) 735 08 60

Jean De Lannoy

Avenue du Roi 202 / Koningslaan 202
 B-1060 Bruxelles / B-1060 Brussel
 Tél. (02) 538 51 69
 Télex 63220 UNBOOK B
 Fax (02) 538 08 41

Document delivery:

Credoc

Rue de la Montagne 34 / Bergstraat 34
 Bte 11 / Bus 11
 B-1000 Bruxelles / B-1000 Brussel
 Tél. (02) 511 69 41
 Fax (02) 513 31 95

DANMARK

J. H. Schultz Information A/S

EF-Publikationer

Ottliavej 18
 DK-2500 Valby
 Tlf. 36 44 22 66
 Fax 36 44 01 41

DEUTSCHLAND

Bundesanzeiger Verlag

Breite Straße
 Postfach 10 80 06
 D-W 5000 Köln 1
 Tel. (02 21) 20 29-0
 Telex ANZEIGER BONN 8 882 595
 Fax 2 02 92 78

GREECE/ΕΛΛΑΔΑ

G.C. Eleftheroudakis SA

International Bookstore
 Nikis Street 4
 GR-10563 Athens
 Tel. (01) 322 63 23
 Telex 219410 ELEF
 Fax 323 98 21

ESPAÑA

Boletín Oficial del Estado

Trafalgar, 29
 E-28071 Madrid
 Tel. (91) 538 22 95
 Fax (91) 538 23 49

Mundi-Prensa Libros, SA

Castelló, 37
 E-28001 Madrid
 Tel. (91) 431 33 99 (Libros)
 431 32 22 (Suscripciones)
 435 36 37 (Dirección)
 Télex 49370-MPLI-E
 Fax (91) 575 39 98

Sucursal:

Librería Internacional AEDOS

Consejo de Ciento, 391
 E-08009 Barcelona
 Tel. (93) 488 34 92
 Fax (93) 487 76 59

Librería de la Generalitat de Catalunya

Rambra dels Estudis, 118 (Palau Moja)
 E-08002 Barcelona
 Tel. (93) 302 68 35
 302 64 62
 Fax (93) 302 12 99

FRANCE

Journal officiel Service des publications des Communautés européennes

26, rue Desaix
 F-75727 Paris Cedex 15
 Tél. (1) 40 58 75 00
 Fax (1) 40 58 75 74

IRELAND

Government Supplies Agency

4-5 Harcourt Road
 Dublin 2
 Tel. (1) 61 31 11
 Fax (1) 78 06 45

ITALIA

Licosa Spa

Via Duca di Calabria, 1/1
 Casella postale 552
 I-50125 Firenze
 Tel. (055) 64 54 15
 Fax 64 12 57
 Telex 570466 LICOSA I

GRAND-DUCHÉ DE LUXEMBOURG

Messageries Paul Kraus

11, rue Christophe Plantin
 L-2339 Luxembourg
 I-50125 Firenze
 Tél. 499 88 88
 Télex 2515
 Fax 499 88 84 44

NEDERLAND

SDU Overheidsinformatie

Externe Fondsen
 Postbus 20014
 2500 EA 's-Gravenhage
 Tel. (070) 37 89 911
 Fax (070) 34 75 778

PORTUGAL

Imprensa Nacional

Casa da Moeda, EP
 Rua D. Francisco Manuel de Melo, 5
 P-1092 Lisboa Codex
 Tel. (01) 69 34 14

Distribuidora de Livros Bertrand, Ld.ª

Grupo Bertrand, SA
 Rua das Terras dos Vales, 4-A
 Apartado 37
 P-2700 Amadora Codex
 Tel. (01) 49 59 050
 Telex 15798 BERDIS
 Fax 49 60 255

UNITED KINGDOM

HMSO Books (PC 16)

HMSO Publications Centre
 51 Nine Elms Lane
 London SW8 5DR
 Tel. (071) 873 2000
 Fax GP3 873 8463
 Telex 29 71 138

ÖSTERREICH

Manz'sche Verlags- und Universitätsbuchhandlung

Kohlmarkt 16
 A-1014 Wien
 Tel. (0222) 531 61-0
 Telex 112 500 BOX A
 Fax (0222) 531 61-39

SUOMI

Akateeminen Kirjakauppa

Keskuskatu 1
 PO Box 128
 SF-00101 Helsinki
 Tel. (0) 121 41
 Fax (0) 121 44 41

NORGE

Narvesen information center

Bertrand Narvesens vei 2
 PO Box 6125 Etterstad
 N-0602 Oslo 6
 Tel. (2) 57 33 00
 Telex 79668 NIC N
 Fax (2) 68 19 01

SVERIGE

BTJ

Tryck Traktorvägen 13
 S-222 60 Lund
 Tel. (046) 18 00 00
 Fax (046) 18 01 25

SCHWEIZ / SUISSE / SVIZZERA

OSEC

Stampfenbachstraße 85
 CH-8035 Zürich
 Tel. (01) 365 54 49
 Fax (01) 365 54 11

ČESKOSLOVENSKO

NIS

Havelkova 22
 13000 Praha 3
 Tel. (02) 235 84 46
 Fax 42-2-264775

MAGYARORSZÁG

Euro-Info-Service B.T.

Rádäy u. 24/B
 H-1092 Budapest
 Tel. (1) 36 1 118
 Fax (1) 36 1 72 83

POLSKA

Business Foundation

ul. Krucza 38/42
 00-512 Warszawa
 Tel. (22) 21 99 93, 628-28-82
 International Fax&Phone
 (0-39) 12-00-77

CYPRUS

Cyprus Chamber of Commerce and Industry

Chamber Building
 38 Grivas Digenis Ave
 3 Deligiorgis Street
 PO Box 1455
 Nicosia
 Tel. (2) 449500/462312
 Fax (2) 458630

TÜRKIYE

Pres Gazete Kitap Dergi Pazarlama Dağıtım Ticaret ve sanayi AŞ

Narlıbahçe Sokak N. 15
 İstanbul-Çağaloğlu
 Tel. (1) 520 92 96 - 528 55 66
 Fax 520 64 57
 Telex 23822 DSVO-TR

CANADA

Renouf Publishing Co. Ltd

Mail orders — Head Office:
 1294 Algoma Road
 Ottawa, Ontario K1B 3W8
 Tel. (613) 741 43 33
 Fax (613) 741 54 39
 Telex 0534783

Ottawa Store:

61 Sparks Street
 Tel. (613) 238 89 85

Toronto Store:

211 Yonge Street
 Tel. (416) 363 31 71

UNITED STATES OF AMERICA

UNIPUB

4611-F Assembly Drive
 Lanham, MD 20706-4391
 Tel. Toll Free (800) 274 4888
 Fax (301) 459 0056

AUSTRALIA

Hunter Publications

58A Gipps Street
 Collingwood
 Victoria 3066

JAPAN

Kinokuniya Company Ltd

17-7 Shinjuku 3-Chome
 Shinjuku-ku
 Tokyo 160-91
 Tel. (03) 3439-0121

Journal Department

PO Box 55 Chitose
 Tokyo 156
 Tel. (03) 3439-0124

RUSSIA

CCEC (Centre for Cooperation with the European Communities)

9, Prospekt 60-let Oktyabria
 117312 Moscow
 Tel. 007 095 135 52 87
 Fax 007 095 420 21 44

ISRAEL

ROY International

PO Box 13056
 41 Mishmar Hayarden Street
 Tel Aviv 69865
 Tel. 00972 3 496 108
 Fax 00972 3 544 60 39

SINGAPORE

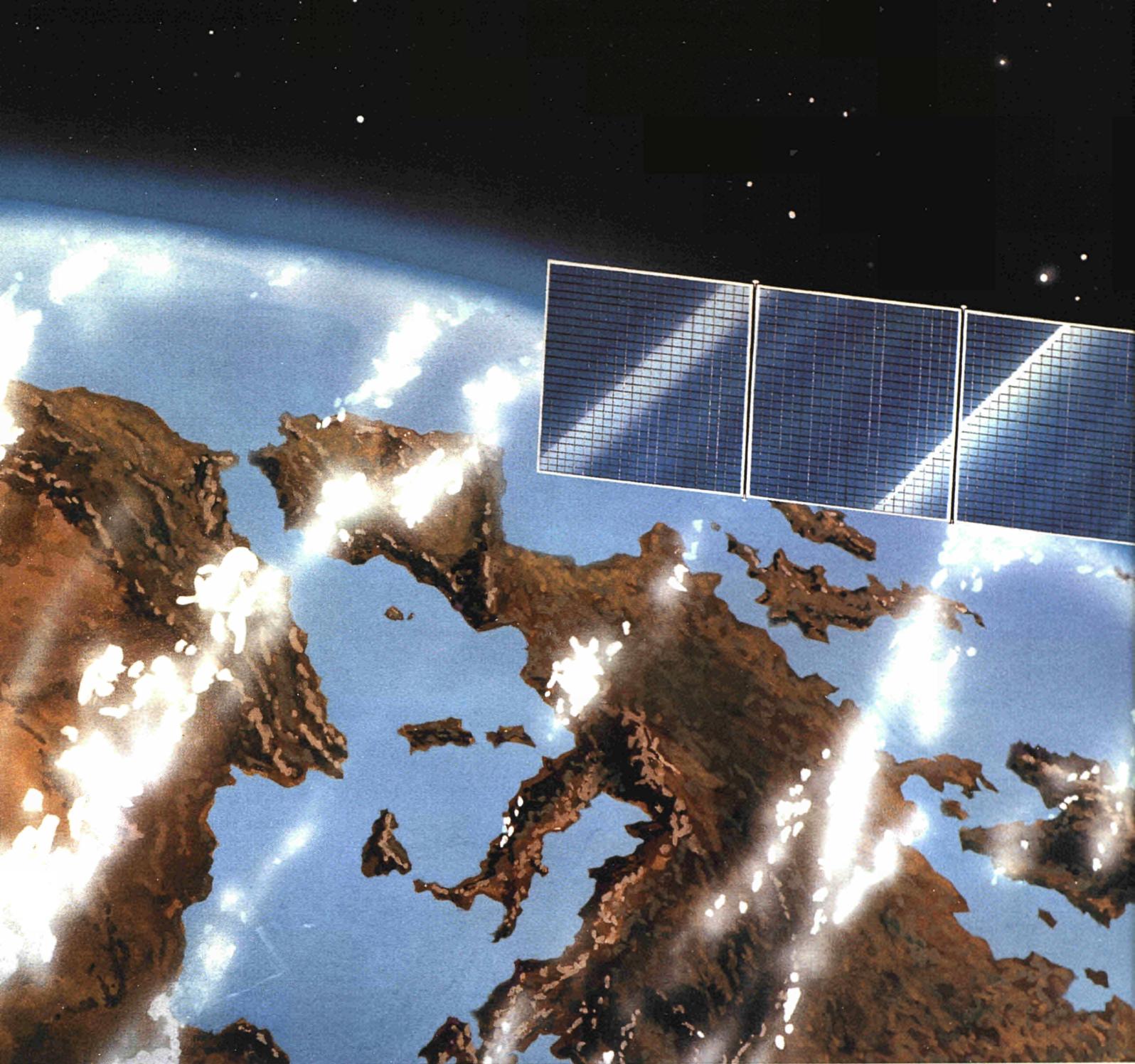
Legal Library Services Ltd

STK Agency
 Robinson Road
 PO Box 1817
 Singapore 9036

AUTRES PAYS
 OTHER COUNTRIES
 ANDERE LÄNDER

Office des publications officielles des Communautés européennes

2, rue Mercier
 L-2985 Luxembourg
 Tél. 499 28 1
 Télex PUBOF LU 1324 b
 Fax 48 85 73/48 68 17



SOCIÉTÉ EUROPÉENNE DES SATELLITES - L-6815 Betzdorf, GD of Luxembourg

