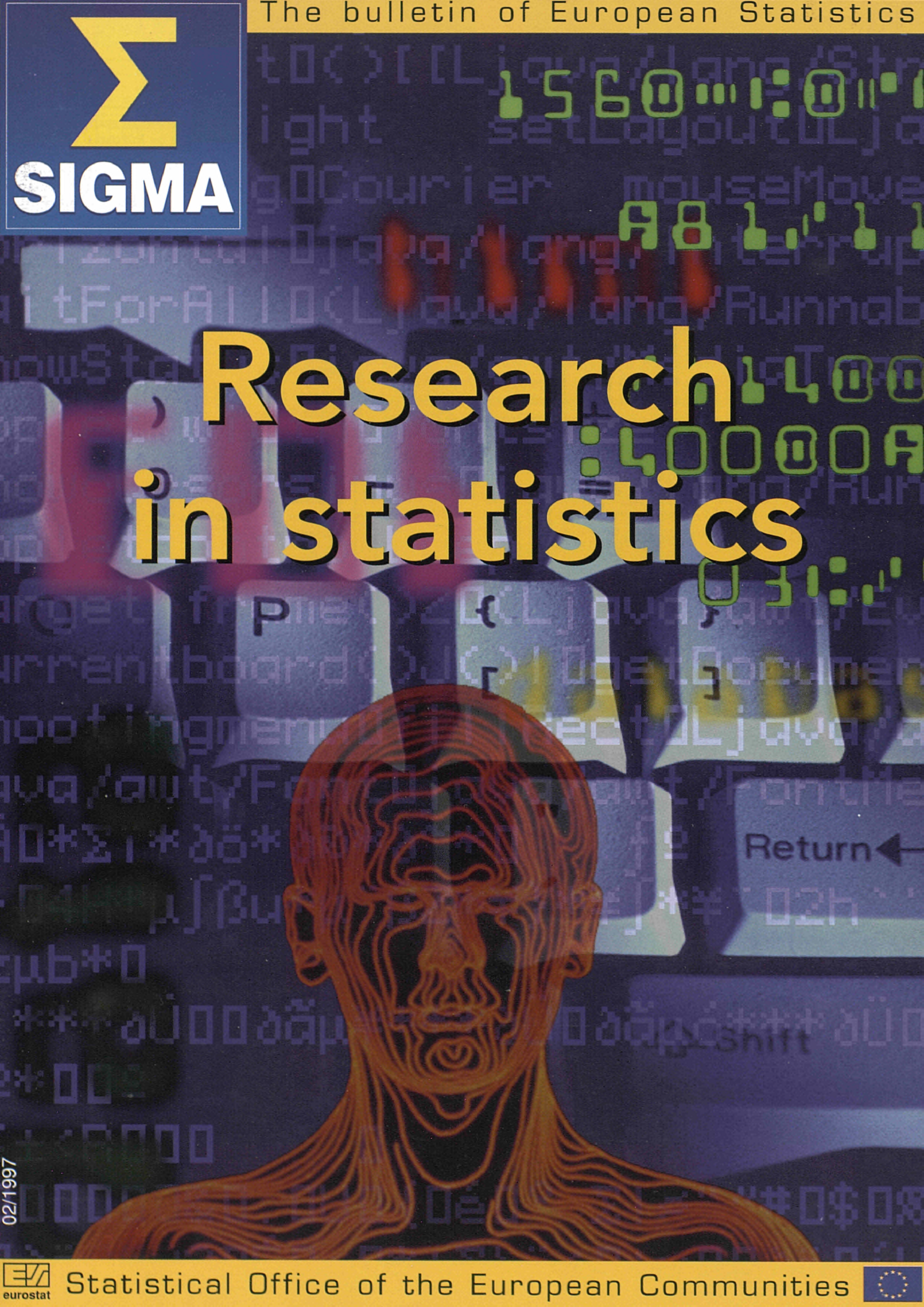
The bulletin of European Statistics

SIGMA

Research in statistics



02/1997

eurostat Statistical Office of the European Communities

Eurostat Data Shops

Statistics on tap

Eurostat Data Shops are now open in five EU cities, Paris, Rome, Milan, Brussels and Luxembourg – and also in New York. And soon the United Kingdom and Spain will have them, too – with others planned in the remaining Member States. The shops enable anyone to plug into a huge range of data on the EU.

The Data Shop service covers supplying simple and complex statistical data, providing access to Eurostat's traditional and electronic publications, and also producing 'made-to-measure' comparative tables and statistical charts – by fax or e-mail or on paper, diskette, magnetic tape or CD-ROM.

For example, the New Cronos database contains more than 70 million items of socio-economic statistical data covering Member States and, in numerous cases, the USA, Japan and the main trading partners of the EU as well. And COMEXT contains statistics on intra- and extra-EU trade for several thousand products.

The shops are designed to serve the man and woman in the street as well as companies, institutions, government departments and universities.

There can be direct communication with an individual member of staff who will deal with your request and whom you can contact by phone or fax if there is a problem.

Customers are charged according to the type of service.

The present Data Shop network is as follows:

Eurostat Data Shop Luxembourg
2, rue Jean Engling
L-1466 Luxembourg
Tel : + 352-43 35 22 51
Fax : + 352-43 35 22 221
AgnesN@eurostat.datashop.lu

Eurostat Data Shop Bruxelles
rue Joseph II, 121
B-1049 Bruxelles
Tel :+ 32-2-29-96 666
Fax : + 32-2-29-50 125
Piera.Calcinaghi@eurostat.cec.be

INSEE INFO SERVICE
Eurostat Data Shop Paris
195, rue de Bercy – Tour Gamma A
F-75582 Paris Cedex 12
Tel : + 33-(0)1-53 17 89 43/45
Fax : + 33-(0)1-53 17 88 22

HAVER ANALYTICS
Eurostat Data Shop New York
60 East 42nd Street – Suite 2424
New York, NY 10165
Tel : + 1-212-986 93 00
Fax : + 1-212 986 58 57
eurodata@haver.com

ISTAT-CENTRO
DI INFORMAZIONE
STATISTICA-SEDE DI MILANO
Eurostat Data Shop Milan
Piazza della Republica 22
I-20124 Milano
Tel : + 39-2-65 95 133/134
Fax : + 39-2-65 30 75

ISTAT-CENTRO
DI INFORMAZIONE
STATISTICA-SEDE DI ROMA
Eurostat Data Shop Rome
Via Cesare Balbo 11/A
I-00184 Roma
Tel : +39-6-46 73 31 05/02
Fax : +39-6-46 73 31 07/01

midasnet

Sigma – **the bulletin of European statistics produced in Luxembourg by Eurostat**

Chief editor
Daniel Byk

Assistant chief editor
Fons Theis

Editorial team
John Wright
Barbara Jakob
Annika Östergren
Steffen Schneider

Assistant
Virginie Benoit

Layout
Claudia Daman
Quentin Masquelier

Cover
Made by Sams – Luxembourg

Published by
Office for Official Publications of the European Communities

Catalogue number
CA-AB-97-002-EN-C

Brussels • Luxembourg 1997

Printed in Luxembourg

Sigma is available free of charge from:

Eurostat
Press & Communications Team
Room B3/079
Jean Monnet Building
L-2920 Luxembourg
Tel: 352 4301 33444 / 33496
Fax: 352 4301 35349

FERNANDO DE ESTEBAN has ceased to be Managing Editor of *Sigma* on leaving Eurostat to become Director of Computing of the European Commission in Luxembourg. It was under his direction that *Sigma* took a new direction with improved presentation and contents. In our next issue we shall carry an interview with him analysing the relationship between statistics and data processing.

CONTENTS

Research – a key focus for European statistics

by Daniel Defays

Certain words have the capacity to arouse interest, curiosity, even enthusiasm. Say the word 'research' and the image that springs to mind is of a healthier and more comfortable future with better services. You are conjuring with the unknown, but this is challenging not worrying. Other words, however, seem smothered in layers of dust and are simply yawn-inducing or mind-numbingly dull. Unfortunately, for some people, 'statistics' is such a word.

And so it is with mixed feelings, wary curiosity or detached interest that you are likely to approach reading something on 'statistical research'.

These words not only promise two very different things, they are also ambiguous. The research and statistics to which they refer are not the fields usually encountered: they have little to do with laboratory work, ethereal or abstract developments, exotic distributions, asymptotic convergences, or sophisticated models. Instead, they are investigative activities that might lead to long-term improvements in the quality of our statistical information systems through the more intensive use of new methods and techniques.

There is nothing trivial about this task. The process required to transform a publication in a scientific journal into a statistical software tool, or an idea into an innovation is a long and painstaking one that calls for discipline and ingenuity, as well as a detailed knowledge of requirements and techniques.

Turn on the radio in the morning and it is no longer unusual to hear the results of an opinion poll conducted a few days previously, or be given statistics on the number of viewers who watched a particular television programme the night before. Forecasts are advanced using new methods (such as neural networks) that draw on complex structures. The presentation of statistics is becoming more eloquent and better-informed. It is trends such as these that we wish to harness and develop so they can benefit the whole sphere of official statistics, and all users in all European countries.

A long road ahead

The road ahead is long. Processing surveys is a laborious task often taking years, and the results are sometimes incomplete, outdated or poorly targeted. Data collection and processing methods are all too often anachronistic. We need to adapt, modernise and innovate.

Most of the time, the word *innovation* is used only in association with the private sector. This is unfortunate. Public authorities and, more specifically, the statistical services are having to face up to new challenges: limited resources, information-providers who sometimes feel hassled by requests for statistics, and increasing demands involved in describing a socio-economic environment ever more complex and fluid.

In this new society, where information flows are becoming faster and faster, data have acquired quite a different value – a market value – as they become a vital ingredient of economic and political activity. On top of all this, decision-making and data collection centres are forever changing and the arsenal of methods and techniques to be employed is constantly expanding.

We need to innovate

To solve these problems we need to innovate, and probably have to pass through a number of stages: finding and developing less expensive and faster data collection methods; a more judicious use of the vast reserves of data produced by the information society; the automation of certain tasks carried out by statisticians; the formalisation and more intensive use of all non-numerical information accompanying or supplementing data; studies of new methods of presenting and displaying our data; and use of networks.

A'knowledge industry' is beginning to emerge. Statisticians will not only have to try and measure its key parameters, but, as experts in describing populations and distributions, they will be called on to play an active role. This is what R&D activities in the field of statistics aim to achieve. They should help ensure that, rather than stand by waiting to see what the future has in store for us, we go out and shape the course of events to make them suit our requirements.

This cannot, of course, be achieved by one operator or one institution alone. Eurostat is endeavouring to use the European statistical system to channel efforts, promote the emergence of new ideas and facilitate exchanges of technology and know-how. To do so, various integrated research programmes have been launched under the umbrella of the Community framework programmes for research and technological development. These have helped unite researchers, national statistical services and the business world in tackling the problems associated with official statistics.

Given the exiguity of the market for certain products, it has been essential to combine our efforts at a European level to reach the critical mass required for meaningful research.

Clear trend emerging

First results of these initiatives have already begun to appear. The DOSES programme (Development of Statistical Expert Systems) has opened up some interesting avenues: the automatic drafting of comments on the basis of digital data; a drastic reduction in time required to process and publish data from certain surveys as a result of more advanced computerisation; the unifying role of metadata etc.

This progress has been followed by new current activities known under the collective title of DOSIS (Development of Statistical Information Systems). First results of this work are expected over the next few months. These projects deal with areas such as collection of data using EDI techniques; development of new types of interface for statistical databases; analysis of symbolic data; extraction of knowledge; and confidentiality. In its document *Towards the Fifth Framework Programme: scientific and technological objectives*, the Commission suggested that activities in the field of statistics be continued. In parallel with Community activities, Member States have also produced some impressive developments in recent years in areas such as natural-language interfaces, geographical information systems and data collection tools.

A clear trend is emerging. Activities that only a few years ago were still regarded as marginal, even frivolous, have now become a key focus for the European statistical system as a whole. Statistical research is an essential component of our work programmes, an instrument for managing change. It needs to be allied with the desire to use its results to improve our working methods, our tools, our environment and with training activities. Only then will it help improve the quality of our output by enabling us to offer a better service to society and individual citizens.

Daniel Defays,

head of Eurostat unit for research and development, methods and data analysis

'Statisticians must make their presence felt'

Community statistics have found a place in European research framework programmes. On the one hand, they benefit from added value at European level in the form of cooperation and risk-sharing between official statistical bodies, the academic world and the private sector. On the other, they are required to sharpen the focus of research in general on the needs of the users, the democratic society and the individual citizen. How does that work? What are the opportunities for Community statistics in the age of telecommunications? What is the meaning of the term 'user-driven'? Sigma's STEFFEN SCHNEIDER *spoke to* PHOTIS NANOPOULOS, *one of Eurostat's Directors, whose responsibilities include statistical information systems and data research and analysis.*

I ask if the role of a statistical institute is to carry out or promote research?

"First of all", says **Nanopoulos**, "it should be understood that Eurostat does not carry out research itself; it concentrates on promoting scientific research. The Office is not a research institute, but encourages specialised research activities for which it would be difficult to find products on the market or investors.

"We do not directly intervene in research on statistics; we confine ourselves to the role of observer. A large part of statistical research is carried out in universities and scientific institutes. Very often we buy it. This was the case, for instance, regarding the Statistical Analysis System (SAS) – a software for processing and storing statistical data – which was the fruit of research conducted outside the field of statistics.

"As regards areas which directly affect our work, we act as promoters to stimulate and focus research in a certain direction. We create a type of framework.

"What's more, we have the fortune and opportunity of being able to use a Community framework programme *(see the articles on pages 7 and 11)*. Through discussions with the Directorates-General of the European Commission responsible for promoting research we define a place for statistics within the programme. An example of this is the DOSIS programme (Development of Statistical Information Systems).

"The focus of these programmes helps bring us to the notice of the bodies most interested ie universities, software and databank businesses and, of course, the national statistical institutes. The latter are now involved to a greater extent, so that the users' point-of-view is expressed, since Community research programmes are generally user-oriented. There are also other advantages, such as Community added value and cooperation between Eurostat and the Member States, not to mention the exchange of ideas between the official statistical bodies, universities and the academic world, and industry.

"We must draw full benefit from this environment, whilst remaining aware that, compared with the budgets for fields such as industry or transport, statistics have only limited resources.

"Activity remains the crucial factor, since statistics will be subject in the future to various pressures. It will, above all, be the technological dimension – technological improvements – that will overcome these problems. We have to realise that this is the age of telecommunications – the world is based on telecommunication, the processing of information, not only to exchange messages but also as a means of working, buy-

ing, selling and even governing. All these human activities are involved.

"At present we are winning the battle to implement the Fourth Framework Programme, whilst preparing for the fifth one. The main thing will be to get involved in a very committed, visible way in order to encourage the protagonists to make a major effort in preparation for the year 2000. Research has to be one of the main bedrocks, one of the pillars for the success of the statistical programme from the year 2000."

What other advantages, I ask, does Europe currently enjoy regarding statistical research and development?

"European research, by its international character, encourages a cultural mix, a coming together of scientific approaches: the pragmatism of some, the more conceptual analyses of others – a mobilisation of means that brings sure advantages. The work on metadata, symbolic data and automatic codification are in the forefront of research."

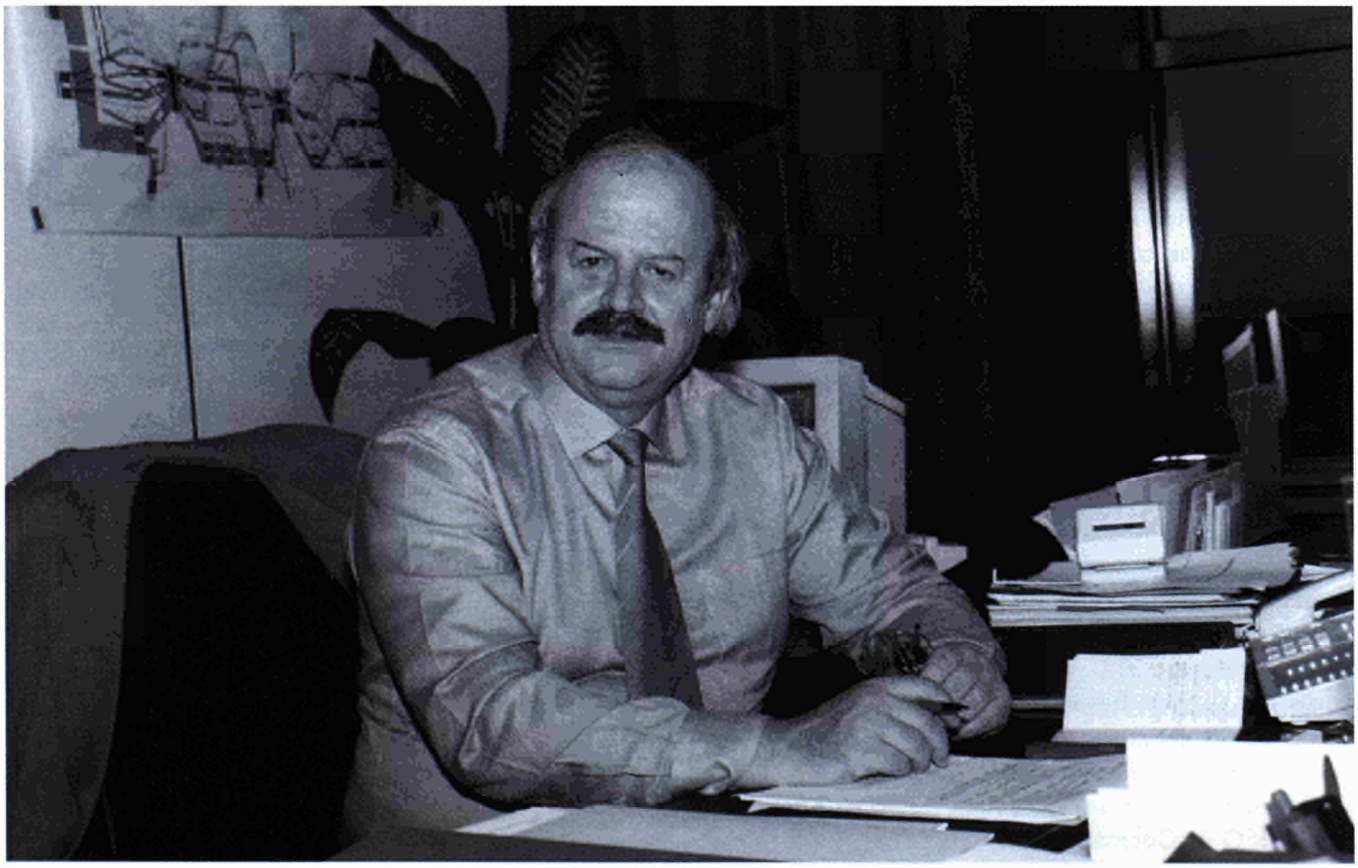
Presence felt

Have statistics always been an obvious component of the framework programme?

Nanopoulos: "Statistics come to mind quite naturally, but they have never been a priority. Statisticians therefore have to make their presence felt to protect it. At present, because of a change in political vision, the Fifth Framework Programme is directed more towards the citizen and the solution of social problems than towards basic research. Research is becoming a genuine instrument for solving social problems. Consequently, in the Fifth Framework Programme, there is greater emphasis on the socio-economic dimension – the service to be provided for citizens and for democracy. And we will succeed in carving out a proper role for ourselves. Even if statistics are not considered a priority, they are and will remain a useful instrument."

Research is, by definition, international, I say. How are relations with non-member countries?

"Several areas of special cooperation have been defined with the USA and Canada and also with Israel. Relations have also been established with the countries of Eastern Europe and these are to be developed in the future. Certain capacities do exist. It would certainly be useful to involve them in the research programmes and to improve their systems. Apart from these, there are also the EFTA countries."



Photis Nanopoulos, Eurostat Director for the statistical information system, research and analysis of data and technical cooperation with Phare and Tacis countries

What is the role of the NSIs? Do they simply express requirements? Do they test and validate systems? To what extent do they themselves carry out research?

"Several institutes lead the way as regards research, for instance the CBS in the Netherlands. They are the leaders in several areas, including even the development of software. In other areas the NSIs are involved as partners in the technological development, assessment and final validation of a system. Virtually all the projects involve several NSIs and almost all the Members States at present participate."

Was this different before?

Nanopoulos: "Yes. At the beginning of the Second Framework Programme and the DOSIS programme a limited number of NSIs participated. The situation changed totally with the Fourth Programme."

What, I want to know, are relations like with industry and the authorities?

"We experience fairly general problems. The tools which will be produced will be of a fairly general nature. If benefits emerge and potential clients appear, industry will be interested in making an investment, since it will be 'sellable'. This pre-supposes that other authorities ie ones which are not necessarily statistical, will have similar needs as regards processing data, validating them and estimating missing data. At present statistical tools exist which have been designed for an authority's own requirements or for an institute which markets the service, such as the IFO in Germany. They participate in the hope of being able to discover aspects of interest to them.

"Some projects are arousing the attention of the private sector and industry because they have market outlets and profits are likely. The collection of data from enterprises is, for instance, an application which attracts anyone developing software for accounting purposes and business management, since it is an additional service which they will try to sell. Using a statistical tool, it is possible at the same time to incorporate national statistical information, to input external information and to integrate it into the business's information, to seek tools to process the software and to carry out one's own analysis. Such an approach arouses the interest of software specialists."

User-driven

The Commission has a research centre in Ispra. What can Eurostat expect from it?

"We have been working in very close cooperation with the Ispra Centre for over ten years. While it is true that the Centre is not directly concerned with statistics, areas of cooperation do exist. In recent years, particularly since the change in Ispra's system, contacts have been intensified. Several projects, such as SUPCOM *(see the article on page 7)*, have been developed. At present we are trying to establish better coordination and to increase consultation."

How, I ask, should a statistical research programme be structured? Should it begin with the technologies and find out what they offer to statistics or should the problems be the starting point?

"Our starting point is requirements", Nanopoulos replies. "The initial analysis is an analysis of demand; it is user-driven. We then research the most appropriate technology – what approach should be adopted to solve the problem. An expression of needs by the users is essential.

"We start with a diagnosis of the information system and an analysis of requirements and on that basis we define a technology project. Two approaches are open to us: if we have a good general description, we take the initiative ourselves, or if we have no good ideas, we are open to the suggestions of others. 'Author-driven' or 'demand-driven' – these are the key terms. The final choice depends on the individual project."

What mechanism is available for passing on the results of research to the operating systems that can be used by NSIs?

"It must be understood that research regarding official statistics is not a theoretical, academic type of research. Most of our research projects are aimed at making optimum use of existing technologies for statistical purposes. It is not a question of developing anything else. Finding out the best use of the Internet or seeking data in each of the Member States are very applied types of research. When such research is concluded, we have a prototype which can be transformed after a development phase into software, a tool that can be used for concrete applications. We are very close to the application and use of research, provided we have sufficient capacity.

"Sometimes", he concludes, "the tools and results are available but insufficient effort has been made to master the technology. The line between development and research is, therefore, difficult to draw."

As well as collecting data on R&D in Europe, Eurostat is also active in research itself – in statistical projects in the framework of the Commission's R&D policy. Sigma's BARBARA JAKOB *asks project manager* JOHN LUDLEY *to tell us more...*

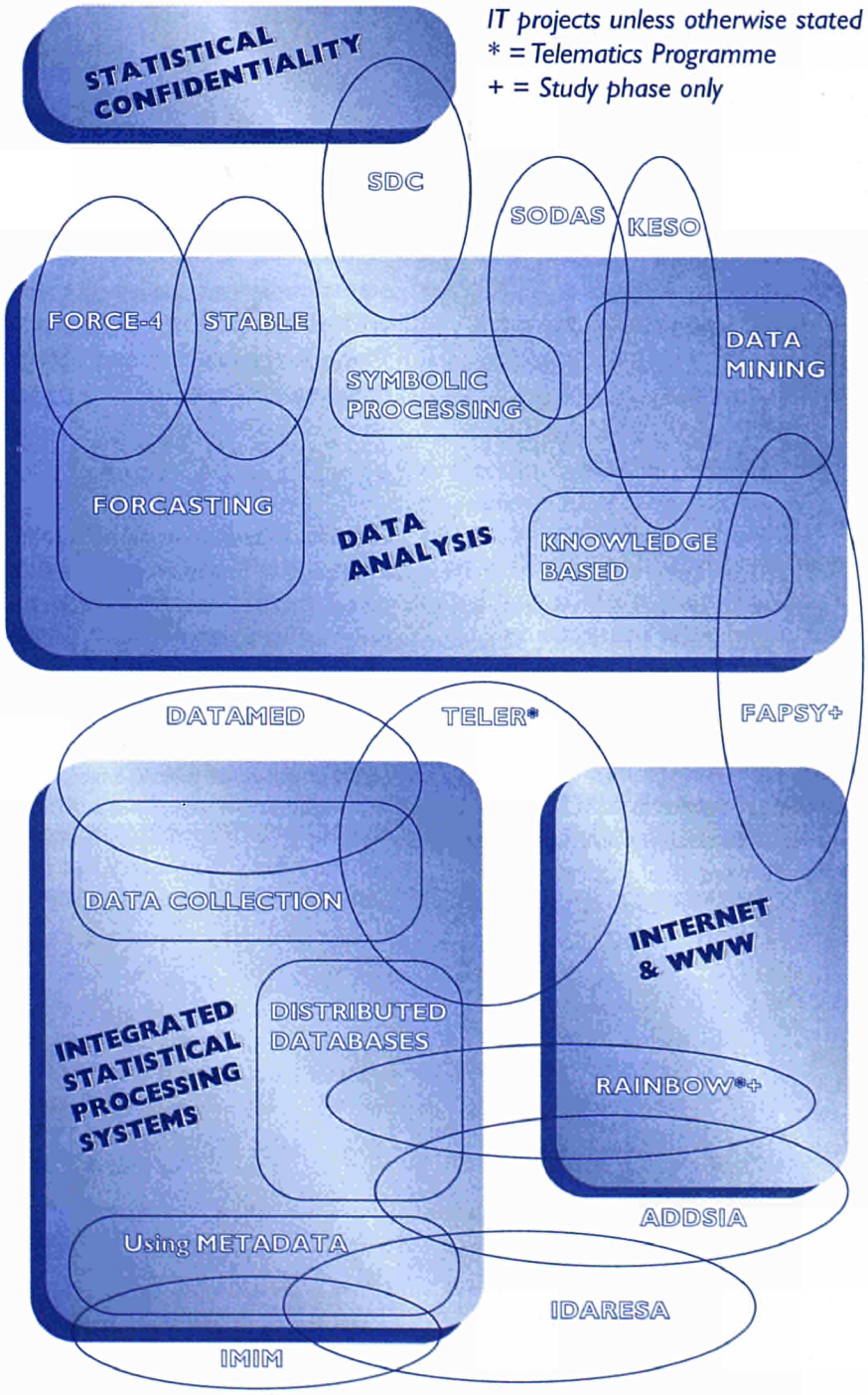
Making the most of statistics through R&D

Research and technological development are recognised as important factors in the competitiveness of the European economies and their position on world markets. The Single European Act, adopted in July 1987, gave the Community a clear mandate to act in this field. General aim of its R&D policy therefore is to strengthen Europe's scientific and technological bases and thus promote employment and quality of life.

This is done through joint research programmes that associate companies, universities and research centres from various European countries. The research themes identified by the Commission and adopted by the Council of Ministers are defined in multiannual framework programmes. These provide basic legal and administrative conditions, scientific and technical targets and content and – last but not least – financial resources.

The current Fourth Framework Programme (1984-1998) has a 13.8 billion ECU budget. It was established to advance the state of knowledge, increase the collaboration and improve the competitive position of European industry and also improve the efficiency of European administrations.

Eurostat's interests in this programme lie in the areas of information technology (DG III), telematics (DG XIII), transport (DG VII) and



This illustrates the complex relationships between the projects and their major themes

Managing statistical R&D projects in Eurostat are (standing, left) Daniel Defays, head of unit, Ralph Jorré, John Ludley, project manager, and (sitting, left) Åsa Jacob, Deo Ramprakash and Tara Byrden

socio-economic issues (DG XII). As in earlier programmes, research projects in official statistics are included.

DOSES of help

Statistics are expected to be up-to-date, relevant and reliable. This in an environment in which Member States are being squeezed on resources for their production and there is increasing reluctance among respondents to provide data. The need to further develop statistical methods and keep them up-to-date is self-evident as a key management tool in both public and private sectors.

"In improving the means of collection and statistical processing – the whole range of 'tools' employed by statistical offices – we are also helping industry", as **John Ludley** puts it.

This was first recognised in 1988 when the DOSES (Development of Statistical Expert Systems) project was accepted for the Commission's general R&D funding. It thus became the Community's first research programme in the field of statistics. "It was quite a narrow area of research", **Ludley** explains. The present programmes are much broader in scope.

Today there are three main areas of R&D managed by Eurostat:

DOSIS (Development of Statistical Information Systems) is the follow-up to the first statistical research programme and is active in information technologies and telematics for administrations. It is oriented particularly towards the needs of producers and users of official statistics.

SUPCOM (Scientific and Technical Support for Community Policies) covers short, usually one-year projects to support the current Eurostat work programme.

A third area of statistical research involves design, use and analysis of statistical indicators in the Targeted Socio-economic Research Programme.

Multi-national cooperation

To emphasise the multi-national cooperative aims of the Fourth Framework Programme, all 12 DOSIS projects involve partners from at least three different countries and, together involve 10 EU national statistical offices. All projects have one or two of three common themes:

- integrated statistical processing
- data analysis, and
- use of Internet and the World Wide Web

with one addressing the more specialised area of

- statistical confidentiality.

Metadata (data descriptive information) is playing a key role in many of the statistical R&D projects. Two projects are particularly concerned with the design and development of metadata-based systems with others strongly dependent on them. Aim of IMIM (Integrated Meta-Information Management) and IDARESA (Integrated Documentation and Retrieval Environment for Statistical Data Aggregates) is "to be able to 'manipulate' statistics by referring to them through their metadata – the information around the data", explains **John Ludley**.

As a result of these projects, a completely-integrated statistical data processing system is expected. This, according to **Ludley**, "will save time and effort and generally reduce the cost of data collection, processing and dissemination."

Reducing costs

In this respect TELER (Telematic for Enterprise Reporting) is extremely interesting for statistical administrations, adds **Ludley**. "It aims to reduce costs of data collection through electronic means. Over three years, standards and software designs are being developed for electronic data collection from enterprises.

"Enterprises will be provided with a PC software system capable of holding enterprise data input, and formating it in a standard way for transmission to NSIs. It will also act as a source of management formation in the enterprises, thereby providing additional benefit.

"Ultimately, if a significant proportion of data can be collected this way for most statistical surveys involving enterprises, that will certainly help NSIs. They will receive data electronically rather than having to key them in. And it will also help the enterprises because they won't have to fill in forms all the time – they can do it automatically through their computing system."

No limit

One big advantage of this project is "it doesn't have to be limited only to statistics. The same technique can be used for sending data from enterprises to tax and social security authorities etc. It may be the beginning of a broader development."

DATAMED (Data Capturing and Interchange in Mediterranean Countries) aims to do a similar job in Italy, Greece and Portugal.

A project just starting is ADDSIA (Access to Distributed Databases for Statistical Information and Analysis). **John Ludley** : "It is about using the WorldWideWeb and Internet for accessing different databases and trying to bring together statistics from a number of sources in a form that could be published or used for research."

In the knowledge-extraction area there is the KESO project for user-oriented information system information services that are looking to the future. **Ludley**: "As organisations – particularly places like hospitals, financial institutions and insurance companies – become more computer-orientated they build up enormous volumes of data. It becomes almost impossible to abstract useful information manually. Aim of this project and others around the world is to be able to use automatic means of looking at these data and finding those that are useful and meaningful – patterns of data you might have not suspected.

"One example: an insurance company might use something like this to detect if certain age groups are a particularly bad risk. It could also be very useful for hospitals. They could find that patients suffering from a particular disease are given treatment that leads to certain side-effects later on.

"The results of many of these projects can be used for different purposes but, of course, our interest is in statistical applications." **Ludley** points out.

'We have to be careful'

One important subject linked to the technological part of statistical production is statistical confidentiality. "The problem of confidentiality is very important to all official statistical offices," **Ludley** says.

"We have to be very careful to avoid releasing aggregated data from which information about individual enterprises could be identified and their data interpreted – the so-called 'Philips effect'. In the Netherlands, Philips is so dominant in the electrical sector, it's difficult to release data in this area without breaching Philips' confidentiality. In fact, the SDC project on tabulating non-disclosive information from masses of sensitive data is led by the Dutch statistical office. It aims to suppress data in the optimum way to minimise loss of useful information while maintaining confidentiality. This needs fairly powerful mathematical techniques."

All these DOSIS' projects will run over a period of two or three

FLAGSHIP OF EUROSTAT'S SCIENTIFIC EFFORTS

Eurostat is launching a journal of scientific collections known as Research in Official Statistics – ROS.

This bi-annual journal has been carefully planned for over 18 months with the support of university professors and other experts from private and public research institutes. It is managed by an editorial advisory committee of mostly academics – Prof Unwin, Prof Lenz and Prof Hand, featured in articles in this issue of Sigma, being members – and an editorial structure within Eurostat.

Other than freely-contributed papers focusing on new techniques and technologies for statistics, ROS covers the main statistical research programmes of Eurostat, notably, DOSIS and SUPCOM.

ROS is available only by subscription but some copies of the first issue are available for review purposes. Further information may be obtained from: Mrs Chantal Sosson (tel: +352 4301-34190).

years with a total budget of about 15 million ECU.

Solving concrete problems

"With DOSIS, we state our interest in a broad range statistical areas and people repond to Calss for Proposals. In the case of SUPCOM we define in fairly precise terms what we require for each project and then issue calls for tender," **Ludley** says.

"Each year, we ask all Eurostat units to define their most pressing R&D tasks, choose those with the highest priority and then issue the call for tender. Generally, we receive around 100 replies. SUPCOM are generally small-scale projects over one year of about 100,000 ECU on average. With a budget of around 3.5 to 4 million a year we can fund up to 30 separate projects.

"SUPCOM is doing useful things for individual units. The work ranges from use of advanced computing techniques to display statistical information, to methods of seasonal adjustment and how best to collect data from household panels.

"It is quite difficult to manage such a large number of small, fragmented projects. This is why we decided from 1997 to group them into a smaller number of larger projects.

"It is probably the statistical research area where you have the closest link to daily practice – the direct application," **Ludley** adds. This is also a primary aim of the Fourth Framework Programme which will be carried through into the Fifth Framework Programme.

Contribution of statistics

"These projects together contribute to achieving the aim of the Commission's Framework Programme – to help the competitiveness of industry. Our contribution is in helping to develop systems which industry can use to lighten their statistical response burden, and also to develop systems to help statistical administrations become more efficient in data collection.

"The main reason it is important to do this at European level is that it is unlikely any individual country could afford such in-depth R&D. In this way every country can benefit from the results."

There are even more advantages. "All these DOSIS projects are being carried out by multinational consortia. This helps the process of communication – working between countries. Within the countries themselves it involves industry, government and the academic community, which adds to the cooperative effect."

[1] *For the complete set of projects see the Eurostat publication* DOSIS Project – technical description *and the Eurostat R&D Web pages.*

THE FOURTH FRAMEWORK PROGRAMME

Activity 1

Research, technological development and demonstration programmes

1. *Information technologies*
2. *Telematics applications*
3. *Advanced communication technologies and services*
4. *Industrial and materials technologies*
5. *Standards, measurements and testing*
6. *Environment and climate*
7. *Marine science and technology*
8. *Biotechnology*
9. *Biomedicine and health*
10. *Agriculture and fisheries*
11. *Non-nuclear energy*
12. *Nuclear fission safety*
13. *Controlled thermonuclear fusion*
14. *Transport*
15. *Targeted socio-economic research*

Activity 2

16. *Cooperation with third countries and international organisations*

Activity 3

17. *Dissemination and optimisation of results*

Activity 4

18. *Training and mobility of researchers*

A framework for statistical solutions

by Deo Ramprakash

If all goes well, the coming years will present further opportunities for research and technological development in statistics at EU level.

Since 1984 EC research and technological development activities (RTD) have been defined and implemented by a series of framework programmes (FPs). The current Fourth RTD Framework Programme expires at the end of 1998. With 1998 already in view, efforts have begun to follow it up with a Fifth RTD Framework Programme. So far there has been recognition of the importance of R&D in statistics in the FP5 proposal. But, together with all its statistical partners, Eurostat needs to demonstrate yet again the value-added of action at EU level.

The relative shift of the centre of economic gravity from Europe to the Pacific rim has been a major spur to the various RTD Framework Programmes – to maintain the competitiveness of European industry, reduce unemployment and generally to maintain European living standards.

Available data suggest there is a strong correlation between research, technological development and innovation, on the one hand, and productivity, growth and job creation, on the other. Innovative industries and firms create more jobs than others. Most jobs have been created in industries that have increased their R&D expenditure the most – the pharmaceutical, aeronautical and agri-foodstuffs industries.

There is also a strong correlation between R&D and qualified human resources and performance levels of growth, productivity and jobs. Investment in R&D and human capital are relatively lower in Europe than in the United States and Japan. The total number of R&D researchers per 1,000 working population in 1993 was 4.7 in the EU; 7.4 and 8.0 respectively in the US and Japan.

Mrs Edith Cresson, EC Commissioner for Research, Education and Training, says: "A large part of the response to these problems lies in improving our level of knowledge and technology, which puts scientific research at the forefront of our search for solutions."

More responsive to change

The discussion in this area was kicked off by the Commission with a series of preliminary guidelines adopted on 10 July 1996. The Commission has adopted the proposal in the working paper *Towards the Fifth Framework Programme: scientific and technological objectives.* (1)

A number of improvements over FP4 have been incorporated in the Fifth Framework Programme. They include concentrating research on a smaller number of key actions, providing for better coordination of research efforts, and making the research programmes more flexible to respond to changing circumstances.

Whereas earlier phases had addressed the particular needs of industry and the Maastricht Treaty's socio-political objectives, FP5 seeks to concentrate on strengthening the research infrastructure itself – specifically medium and large-scale facilities, networks and centres of excellence.

Six specific programmes are proposed, three thematic or sectoral and three horizontal (1). They are:

- Unlocking the resources of the living world and the ecosystem
- Creating a user-friendly information society
- Promoting competitive and sustainable growth

- Confirming the international role of European research
- Innovation and participation of small and medium enterprises
- Improving human capital.

European statistics in the frame

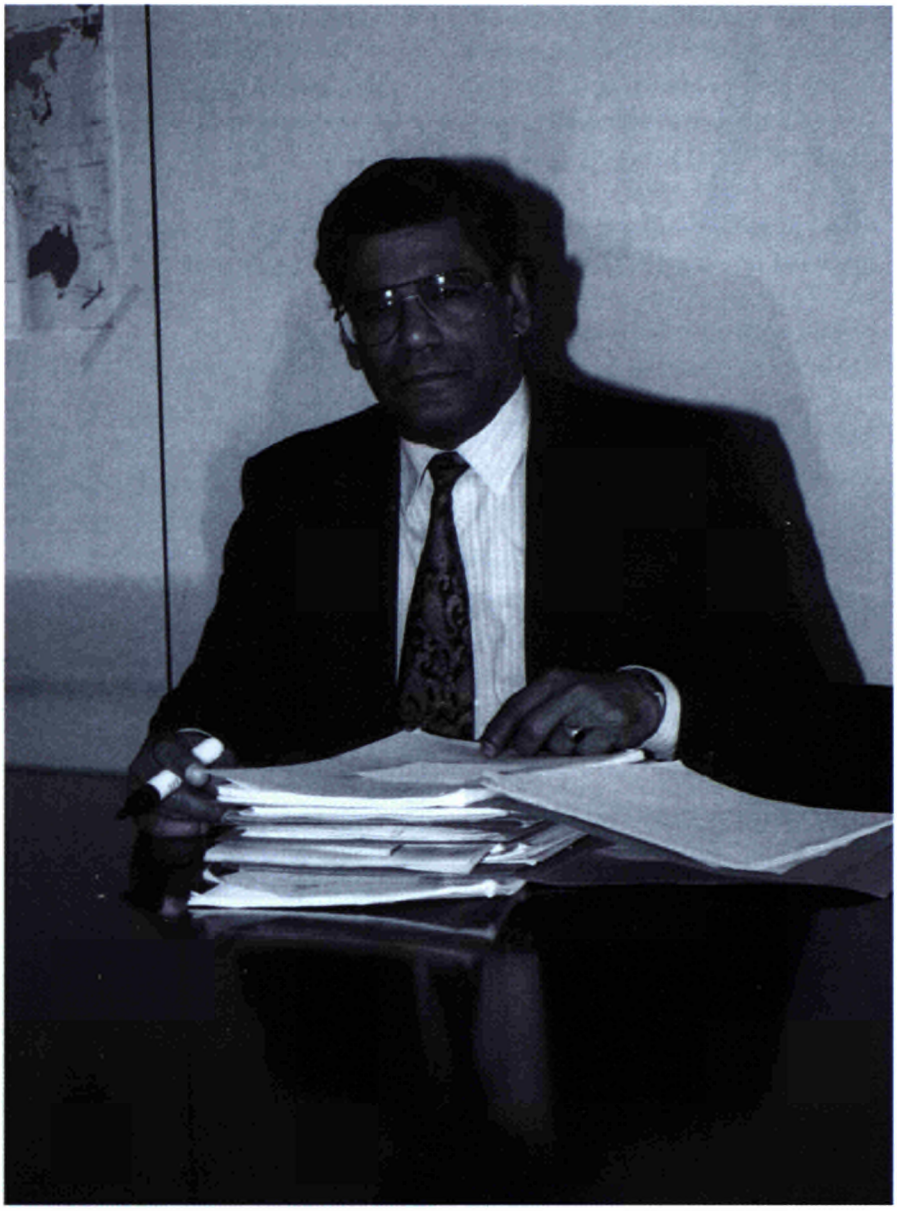
"R&D in official statistics constitutes an important activity for the future development of the European statistical system within the information society", says Photis Nanopoulos, a Eurostat Director, in *Research in official statistics.(2)*

Within increasingly complex and dynamic economic and social realities, statistics would lose their relevance and be unable to contribute to the broader objectives of the framework programmes unless statistical tools, systems and practices were constantly updated and modernised.

The first investment in this area was the DOSES programme which ran from 1989 to 1993. The present R&D programme, DOSIS (Development of Statistical Information Systems), undertaken under the *aegis* of FP4, started in 1994 and will end next year. Its major themes are:

- Information technologies and telematics (Internet and WWW)
- Data analysis
- Integrated statistical processing systems
- Statistical confidentiality.

A major and successful objective of these programmes has been to encourage cooperation through multinational consortia drawn from the academic world, government and the private sector in order to build a long-term, coherent programme for statistical research and development. (3)



Deo Ramprakash is a member of the DOSIS – research in statistics team – of Eurostat's R&D unit

Role of Eurostat

The decision proposed by the Commission includes the following references to statistics:

Information society, key action 4: technology and essential infrastructure-"...technologies and engineering for software and systems, including high-quality statistics"

Support for research infrastructure, key action 5: scientific and technological policy in Europe – "... to be achieved by...the development of a **system of statistics and scientific, technological and innovation indicators.**"

Statistical projects can in principle be submitted not only under these two key actions but, indeed, under other themes and programmes which carry implications for the statistical system. But these possibilities will be explored in the internal working groups.

In any event, for projects to qualify they must contain an R&D content

rather than concern applications exclusively. This R&D component could be the harnessing of technology and techniques to strengthen the statistical infrastructure (or process) in a generic way, i.e. to benefit a number of subject domains simultaneously – for example, the project on Interchange of Data between Administrations. Or the R&D component could be in a particular thematic area – such as the establishment of a Nutrition Habits Statistical Information System under the FP5 programme *Unlocking the resources of the living world and the ecosystem.* There should also be ample scope in FP5 for subsuming developments of large-scale statistical databases.

There are three dates in the FP5 timetable that are critical:

- June 1997 for the first meetings of the internal working groups
- The approximately March-May 1998 date for the Council Decisions on FP5 and the specific programmes, containing more detailed work programmes
- End-1998 for the calls to tender.

Eurostat has plans to keep the ball rolling:

- Another meeting of its consultative DOSIS working group this September
- An information day at the end of this year at which all researchers in the field of official statistics would be invited to articulate their special interests and expertise in specific areas and so encourage the formation of international consortia
- This will be followed by national information parks.

The basic aim is to have by the beginning of next year the main research avenues and themes that could be supported under FP5.

Networking: the 'actors'

The different 'actors' (or 'stakeholders') are:

- the whole of Eurostat, as both producer and customer of the fruits of statistical research
- the national statistical institutes through the Statistical Programme Committee and DOSIS working group
- the research and academic communities
- the authorities responsible for R&D policy at Commission level
- the European Parliament whose influence is clearly important
- the Council of Ministers.

EU value-added (2)

Under this heading are to

- identify: good practice
  centres of competence
  centres of excellence
- diffuse examples of good practices
- help to promote a culture of multi-disciplinary research in statistics
- set up involvement, partnerships and networking between all actors
- contribute to the development of technology and statistical methodology through shared risks and investment
- enhance training and mobility.

Eurostat had organised a meeting of national statistical institutes and other experts in the spring to discuss organisational structures, principles that should underlie the work programme, and an implementation plan.

In addition, with the aim of promoting technology transfer and the diffusion of good practices, Eurostat is visiting national statistical institutes to

- create an inventory of R&D statistical activities
- identify needs in this area
- use this information to promote technology transfer.

If all goes well in the next few months, the FP5 will present further opportunities for research and technological development (RTD) generally, and, within that, for RTD in statistics. This should help the Union to hold its own in an environment of intensifying globalisation and competition. Multi-disciplinary teamwork and partnerships between all players are essential for progress.

But, above all, the emphasis in the programme on the conversion of research into application should know no boundaries.

Notes

(1) COM(97) 142 dated 25 April 1997.

(2) Arguments at the level of statistical components were given in the papers by Mr Photis Nanopoulos presented to the ILIS seminar, Olympia, 1995, and the ISI/BEA seminar, Washington, 1996.

(3) Individual DOSIS projects are outlined in the Eurostat publication DOSIS project, technical description, ISBN 92-827-8728-1.

In this article DR JEAN LOUIS ROOS *from INSEE discusses computerised generation of diagnoses in natural language for statisticians. Or put more simply...*

Darwin and the Alien

A German industrialist may need information on his sector of activity in other European countries. A local residents' association may require immediate and day-to-day analysis of demographic and social data on its particular area. A politician may prefer an analysis of economic conditions in the language of a competent economist – rather than a mass of incomprehensible figures...

For all these examples, figures already exist, but commentary and analysis are costly and time-consuming, if not simply non-existent. Computerised generation of commentaries is the only way to minimise costs in this area.

Generalisation of computer use for processing information, development of major database management systems and dominance of computers and networks – these have transformed the work of statisticians and economists. Information is no longer a scarce resource to be studied in great detail by specialists. It has become so abundant we are faced with a relative scarcity of statisticians for its analysis.

Computer applications are obviously proliferating in statistics. They are certainly becoming more efficient. But nothing can replace language when it comes to pertinent analyses. The next few years will be a phase of systematic introduction of software capable of carrying out the analytical work of statisticians and economists.

Use of such tools will enable the generation of natural-language diagnosis of the contents of any database. This will also certainly be transmitted by voice systems. Development of the Internet means we cannot do without such systems. 'Surfers' will become increasingly interested in information on specific subjects that is not only fresh and objective but delivered in a language they understand. This means their language.

INSEE responds

In response to my proposals, INSEE started taking an interest in work on computerised generation of commentaries as long as seven years ago. First studies concentrated on analysis of the French economy and, more particularly, interpretation of economic survey data. Construction of several models and software products was followed by a prototype. This produced acceptable results, although its application remained particularly complex.

Eurostat is interested in such diagnostic systems for two reasons. On one hand, Eurostat loads statistical data into large databases but hasn't the time to analyse them all.

A diagnosis-generation system is based on two types of information: figures found in a database and the semantics of those data. In practice, the semantics can be broken into two groups: a general part applicable to all statistical data (eg the word increase can be used when a variation is positive), and a specific part applicable to the series being processed. Coding this specific semantic system is particularly burdensome and fraught with problems.

On the other, in a Community that now has 15 Members, a diagnosis in a specific statistical area often must be presented in several languages.

INSEE's use of dictionaries in the system it developed enabled it to produce natural-language text in other languages. And in the meantime in Germany, the IFO (Institute of Economic Research) in Munich constructed a system of economic diagnosis in German.

Eurostat was convinced that cooperation between the two institutions was best way forward. This was done under the DOSES programme. Hence construction by INSEE and IFO, over two years, of an experimental system with capacity to generate natural-language text in French and German. A more modest version in English was tested in parallel. Results were positive. But progress to industrialisation of the software depended on availability of more financial resources. These Europe was ready to provide.

Just a few seconds for 100 pages

In spite of its limitations, the diagnostic system is already giving evidence of its rich potential. Statisticians can use it as a'non-human assistant' to generate correctly-drafted text analysing information in a database. This 'assistant' is capable of making sure no important elements are omitted. Its analyses are, of course, totally independent of subjective external factors.

The software can produce large quantities of text on hundreds of series, in next to no time. Just a few seconds for a hundred pages!

INSEE tests with statisticians have shown it was very difficult to tell the difference between computer-generated...

... and human text

The system can serve as a reference. Once operational, it incorporates 'good practice' – what must be done, said and written in a given situation. This preserves consistency of judgements and analyses, and choices made in the diagnosis can be explained.

Darwin helps Alien

The system is called Alien (logical assistant for the expert interpretation of numerical data).

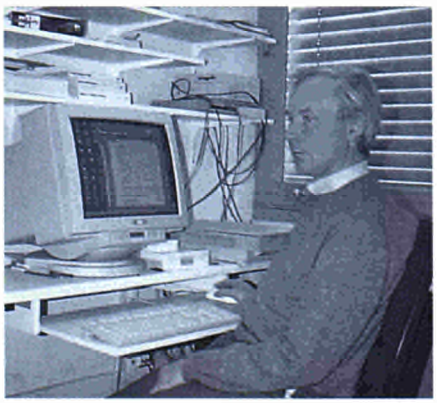
Alien was still difficult to encode. Experience showed users were not content with diagnostic texts alone. They also wanted graphics, tables, pagination and a whole range of user-friendly management tools. To meet these requirements, Eurostat financed development of a new product called Darwin (automatic diagnostics and report under Windows), intended to provide a complete working environment under Windows.

This is now operational and has been handed to Eurostat. There have been several test applications in the areas of transport, trade and industry, but it has not yet entered production. Darwin can be used not only for generating text under Alien but also for structural analysis of knowledge, its modification, interactive creation of graphics and pagination under WinWord. It can also be used, by simply changing dictionaries, to modify the generation language (albeit so far keeping to French'grammar').

Future prospects

One technical advance of Alien included in Darwin is its capacity for diagnoses on relatively complex series. It can identify absence of certain series and write *total rail traffic (except Denmark, Italy and Luxembourg)* when the series for these countries are unavailable. It can also distinguish between estimated values and definitive values and quote them.

INSEE tests with statisticians have shown it was very difficult to tell the difference between computer-generated and human text. So it's expected that soon Alien could be used for production. The



Dr Jean Louis Roos from INSEE is working on a Commission statistical project

operation will consist in circulating, to the 2,000 enterprises providing data for the monthly survey of industrial production, a letter containing a commentary on the situation in their own sector of activity.

Another project, far less advanced but much more ambitious, envisages transmission to France's 36,000 mayors of a report containing an analysis of results of the next population census as applied to their municipalities. This is a particularly complex task because Alien cannot be used. It will be necessary to design a new software system with capacity to manage particular data (results of censuses) and, above all, capable of analysing these results in relation to specific references and breakdowns.

So this future software should be able, by studying various age groups, to write the following type of text: *Average age of the population of municipality X is higher than the average for the Department but has fallen considerably since the last census.*

This is a fascinating operation. Unfortunately, resources earmarked for it are in inverse proportion to importance of the objectives. And time is of the essence. So it's far from certain that this software will reach the operational stage.

But, despite all the problems, full-scale production of computer-assisted diagnosis is getting closer. Soon it should be possible, with the help of Darwin and subject to a little more development work, to carry out computer-assisted generation of yearbooks, commentaries for learned journals and comments associated with databases on the Internet – and do so in several languages.

Note: *The Darwin product, entirely written in C, is the property of Eurostat. The CIA software, which is the basic symbolic interpreter of the system and used for the generation process, is the property of the author.*

Alien is based on the following principles: The system does not process statistical series. It processes indicators that are objects with a common form. These objects contain the semantics of the series, either directly or as inherited traits. All vocabulary required for text generation is placed in external dictionaries. Generation of a diagnosis is in three stages – each lasting only a few seconds or fractions of seconds:

1. *Construction of indicators, description of their semantics and definition of the links between them in the form of an n-step tree (from 0 to infinite)*

2. *Analysis of all situations described in the mathematical formulae that form a part of the semantics. This analysis is based on the values of the series (level, variations, accelerations) and triggers calculation of 'judgements' based on observation of the past*

3. *Sorting through this information and retaining only the most pertinent.*

These three stages are independent of the language. At the end of the third, Alien 'knows' what it must generate. It doesn't yet know 'how' to say it. The generation phase that follows calls up a phrase construction tool and the dictionaries. Modification of the language of the dictionaries modifies the language used for generating the text – but not the grammatical rules.

This is an example of changes in dictionaries (English and Italian dictionaries are at present simple machine translations of French terms and need to be improved):

Dans l'Union européenne , en 1994, le trafic ferroviaire total (national et international) (excepté Danemark. Italie, Luxembourg) est resté faible (644 035.5 milliers de tonnes), mais a augmenté très sensiblement (3.5%). Cette augmentation a été observée pour la Belgique (10.2%), pour l'Allemagne (4.8%), pour l'Espagne (9.9%),pour la France (5.9%) et pour les Pays-Bas (6.5%); à l'inverse, principalement pour la Grèce (-58.0%) et pour le Portugal (-1.3%), le trafic ferroviaire total (national et international) a baissé.

Nell'Unione Europeo nel 1994, il trafico ferroviare totale (nationale e internationale) (eccetto Danimarca, Italia, Lussemburgo) è restato debole (644 035.5 migliaio de tonnellata) ma ha aumentato fortissimo (3.5%). Questa aumento è stata osservata per il Belgio (10.2%), per la Germainia (4.8%), per la Spagna (9.9%), per la Francia (5.9%) e per I Paesi Bassi (6.5%); all'inverse soprattutto per la Grecia (-58.0%) e per il Portogallo (-1.3%) il trafico ferroviaro totale (nationale e internationale) ha diminuato.

In the European Union in 1994, the total traffic by railways (national and international) (except Denmark, Italy, Luxembourg) is resided (644 035.5 thousand tonnes) but has increased very clearly (3.5%). This increasing has been noted for Belgium (10.2%), for Germany (4.8%), for Spain (9.9%), for France (5.9%) and for the Netherlands (6.5%); on the other hand particularly for Greece (-58.0%) and for Portugal (-1.3%) the total traffic by railways (national and international) has decreased.

Paperless revolution...

... at CBS, Netherlands

The Netherlands Government wants to make big reductions in the administrative workload of industry. Business surveys by the national statistical office (CBS) are a small part. CBS can therefore contribute to the reductions via electronic interchange of data with industry.

Intensive use is already made of the new technologies – not only for data collection. Statistical information lends itself extremely well to electronic dissemination. CBS therefore intends to offer its users the possibility of consulting the enormous volume of statistical information that it has at its disposal via the electronic media.

Survey workload

CBS has been asked to reduce its survey workload by at least 100,000 hours. That figure represents a reduction of more than 10%. It will have to be achieved mainly in those areas that impose the heaviest statistical workloads on industry, such as statistics on foreign trade, industrial production and employment and wages. Use of registers kept by other bodies and collection of data by of EDI (electronic data interchange) will both play an important part. CBS also intends to deploy account managers and outworkers to reduce the workload via better coordination with its information-providers.

CBS will gauge the effects of these measures using a survey workload indicator and will periodically report on its findings to industry and government. In addition, it has been given the task of examining whether the creation of a coordination and reporting centre for government statistical surveys would be useful in avoiding situations in which government bodies ask for the same information more than once.

Corporate Tax Information System

One external register CBS will be using is a data file kept by the tax authorities, the Corporate Tax Information System. This system contains, among other things, the annual fiscal accounts of Dutch businesses. These annual data are supplied to the tax authorities by businesses as part of their tax returns. If CBS also makes use of this register, it will be possible to cut by two-thirds the size of the sample (currently 8,000) used in a fairly burdensome annual survey, and reduce the survey workload by more than 7,000 hours. Unfortunately, data for other large businesses cannot be retrieved from the same system, although the survey workload imposed on them will also decline – thanks to use of EDI.

CBS module in wage records

Changes in employment and wage statistics will bring about a considerable reduction in the survey workload. The structure of these statistics underwent fundamental revision in 1995. It now largely matches definitions used by industry's wages departments. This reduced the survey workload at a stroke and permitted transition to EDI, the introduction of which has been underway across the board since 1995. To this end, a special information bulletin was sent to several tens of thousands of information-providers at the beginning of 1995.

In addition, various computer service companies, managerial firms and suppliers of wage-management programs were approached. As a result, nearly all the information-providers have incorporated a special CBS module in their wage management system.

Electronic transmission by municipalities

In 1995 nearly 170 municipalities began transmitting financial data to CBS electronically. Detailed data used to compile 30 items of statistics are now collected six times a year. CBS hopes all municipalities will adopt EDI in future – not only because it will reduce the statistical workload but also enhance the reliability of statistical information and make the results available more quickly.

All municipalities have automated their population registers. As a result CBS now receives population data electronically rather than on paper. This has meant a major adjustment. The transition, which involved solving a number of running-in problems, was made in 1995. Production of statistics is expected to accelerate in the coming years. Options for compiling more detailed population statistics and using them in other CBS statistics will also increase considerably.

Coming soon: 'combined electronic questionnaire'

EDI developments should lead eventually to a number of 'combined electronic questionnaires'. Each information-provider will need to complete just one questionnaire for each government department to discharge its statistical obligations. CBS has launched a pilot project designed to solve associated problems. Results of the project should be apparent after the year 2000.

What are metadata? Why are they important? Where are they at? ANNIKA ÖSTERGREN *finds some answers in Sweden...*

A place in the sun for metadata

Think of a table of figures. Just figures. To understand its contents, obviously we need more information. If its title is *unemployment* we have one pointer. Add dates, say *Jan 1997* and *Dec 1996*, and the focus becomes sharper.

But what exactly do the data refer to? What definition of unemployment is used? How was the survey conducted? How many people were interviewed? What was the non-response rate?

This sort of information on data is what statisticians call *metadata*. Without these extra ingredients raw data are of little use to anyone other than the person who compiled the table.

According to Professor Bo Sundgren of Statistics Sweden (SCB), metadata can be defined as *all the information used to explain data so the user can interpret them correctly.*

Metadata must catch up

Metadata were left behind when computers took off. As we all know, computers are extremely good at processing numbers, but less so at handling text. While data-processing has been automated, metadata have been sidelined by developments.

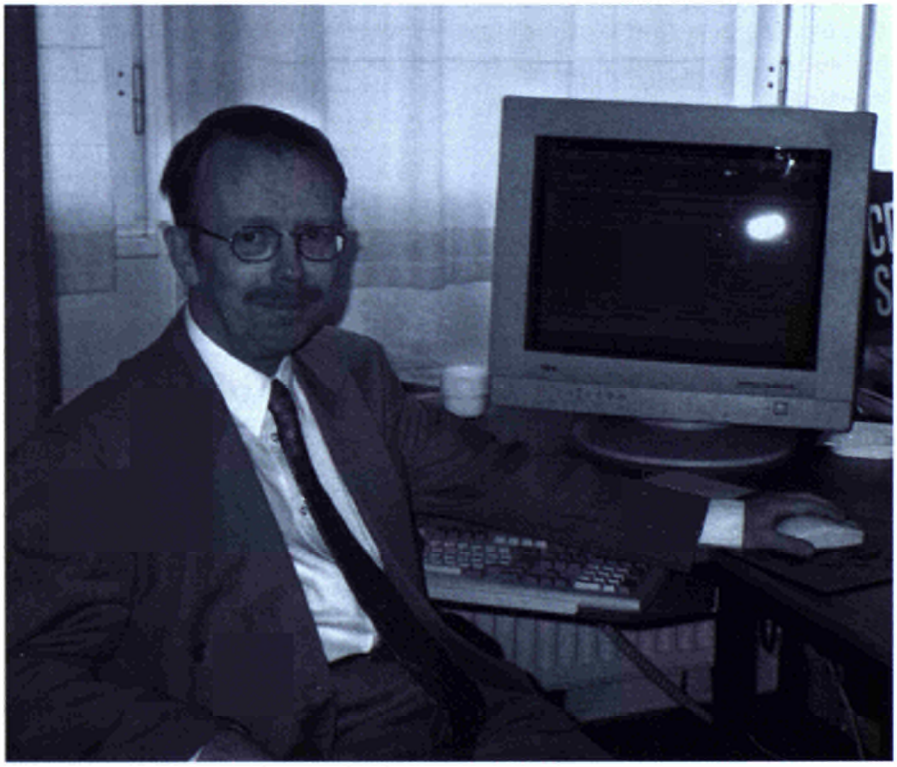
In fact, metadata have always been around but were not seen as important when all data were presented in print. Since data became available electronically, they have become more and more detached from those who collect, compile and process them. This is why it is very important that data are documented so that everyone can use them intelligently.

As Professor Sundgren puts it, statisticians had all the meta-information in their heads and, if anyone wanted further details about a survey, they had to ask the statistician. This is highly impractical for the user. Linking data to metadata is a way of systematising information so everyone can share it.

Template for all surveys

For some years, Statistics Sweden has sought systematically to document information on data. Since the beginning of this year, when all official Swedish statistics were loaded on to the Internet (home page: *http://www.scb.se*), users have also had access to meta-information via SCB's databases.

At best, the information covers everything from administrative data – such as who is responsible for a survey and what it costs – to data on sampling, how the survey was run, variables used and how data were processed. At worst, it gives only definitions of



Professor Bo Sundgren is Head of Statistical Informatics – Department of Research and Development at Statistics Sweden

variables and names of contact people. Ultimate aim is for statistical products to be accompanied by detailed documentation.

Professor Sundgren and his colleagues have developed a template for completion by statisticians so each survey is documented in the same way. At present, there are only three products documented in great detail. Documentation for one – labour force surveys – runs to some 150 pages. Other products have descriptions of around 10 to 15 pages.

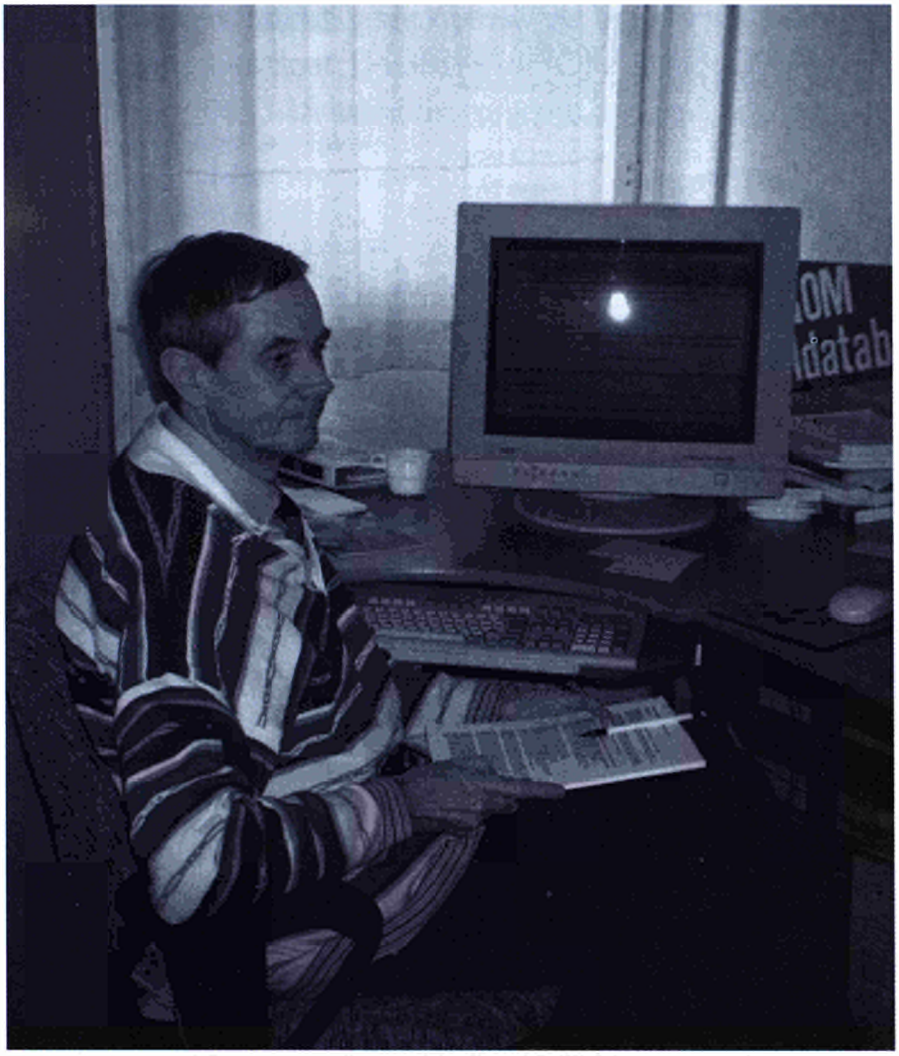
SCB databases even display some important additional information, such as breaks in time-series, directly on-screen when data are downloaded.

Hand-in-hand

To ensure data and metadata go hand-in-hand, provision of documentation should become a reflex action by statisticians. But according to Professor Sundgren there has been a mixed response by statisticians to this idea.

In his view, some perhaps see it as a threat to their position as experts; others as a source of embarrassment, as it might reveal mistakes and shortcomings in the surveys. But most understand that for end-users to use data correctly it is important to have access to *all* information.

The user, for example, will be able to examine the assumptions made in a survey. Using other parameters, results might have been different. Professor Sundgren also feels it is important for users to know how statistical measurements are defined. Without such knowledge, it is difficult to discuss if the measurement used really measures what it's supposed to measure.



Erik Malmborg, senior systems analyst, keeping track of metadata in SCB's databases on the internet. He works with Professor Sundgren

Start to finish

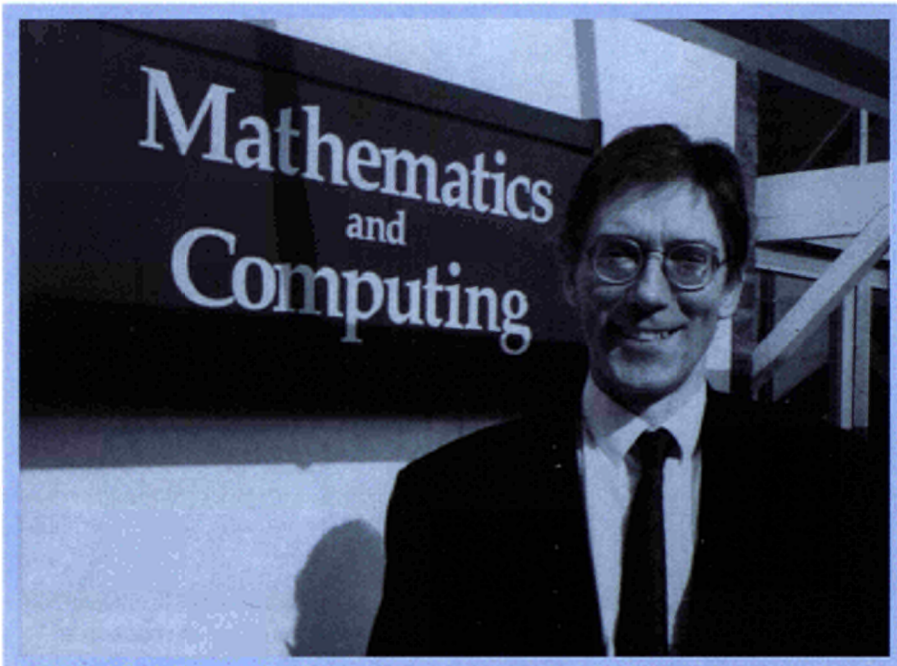
If data are to be well documented and systematically presented, it is important to ensure metadata accompany them right from data collection to final use. Research and development is being carried out in this field at national, UN and EU levels. One project that forms part of the EU's framework programme is IMIM (Integrated Meta-Information Management) headed by Professor Sundgren. It aims to examine how metadata should be integrated to become automatic to the whole statistical process.

One problem of a more technical nature is that current computer programs are rarely able to cope with metadata. SCB's databases are open, which means the user can download data and process them in Excel or some other program. But metadata are seldom transferred.

Professor Sungren feels it is a major shortcoming that commercial programs cannot handle metadata. Metadata should not disappear at the very last stage of the process when the user actually needs them. But he does not think it is the job of national statistical institutes to develop better software. This should be undertaken by the companies already developing programs.

When they heard we were focusing on statistical R&D they said,'You must talk to DAVID HAND'. Sigma's JOHN WRIGHT *took their advice and records the thoughts of an academic who lives and breathes statistical R&D. Not for its own sake but to solve real problems.*

Giving the future a Hand



Professor Hand has been at the Open University since 1988. Before that he spent ten years at London University's Institute of Psychiatry – first as a statistician, then as a lecturer and senior lecturer

***Professor Hand** has done a lot of work for Eurostat over the years – and still advises in various areas. "As more and more data become available thanks to computer and electronic instrumentation, point-of-sale systems, astronomical signals, continuous monitoring of manufacturing processes, and so on, there is a growing need for intelligence embodied in the analysis tool. I was on the 4 million ECU DOSES project to develop such systems – assessing grant applications. Eurostat funded collaborative research throughout Europe on different projects to develop statistical expert systems. I advised them on which projects to fund and went to various European locations to assess progress of the research teams."*

In the old days, says David Hand, your statistician had an image rather like an 18th Century clerk scratching away at tables of numbers. Now he or she is an explorer – with the computer offering infinite new territory unimaginable only a few years ago. Hand should know. He's Professor of Statistics at the Open University in the futuristic city of Milton Keynes, just north of London.

So we discuss what this statistician/explorer gets up to – and why it's so exciting...

Says **Hand**: "I have a lot of research interests, both methodological and applications. The former includes statistical computing, multivariate statistics, foundations of statistics etc. Applications include medicine, psychology and finance – many links with pharmaceutical companies and banks who support collaborative research.

"My projects tend to be about solving real problems. In my view, that's what statistics is for. I am not too interested in abstract mathematics for its own sake.

"So you start with a problem, collect the relevant data and analyse the data using statistical techniques to seek a structure, a pattern and answers to your questions.

"One reason I became a statistician in the first place was because it impacts on just about every area you can think of."

So what, I ask, is the current 'big issue' in statistical R&D?

"If you look at the history of statistics (up to about 1960) you find two driving forces: one is the intrinsic interest in the mathematics of statistics and the other, and more important, problems arising from application areas.

"In more recent history a very important third factor is having a major impact – the power of the computer. It is simply revolutionising the practice of statistics.

"It is this that has changed it from the clerk scratching away at a table of numbers to the modern, most exciting of sciences in which powerful programs are used to probe masses of data."

He continues: "In the early days the main application areas were social and economic statistics. More recently, agricultural statistics have had a tremendous impact, leading to the creation of whole areas such as experimental design. More recently still, medicine has been a driving force for new techniques – things like survival analysis.

"Now I believe financial statistics are one of the big new areas that will lead to the development of entirely new classes of statistical techniques.

"Once techniques have been developed they tend to be applied to other areas. Look at psychology: you find classes of statistical techniques to solve particular problems that are used mainly there but also permeate outside – factor analysis is a good example. Linear structural relation models are another.

"In finance there are traditional areas of statistical applications – actuarial work in insurance, for example. But there are two areas now beginning to have a big impact.

"One has already hit the media – financial markets: futures, options, derivatives etc. That's more probability than statistics and, at least in myth, you have rooms full of boffins developing incredibly refined and mathematically subtle financial instruments. That started some 10 to 20 years ago and has reached quite a peak of sophistication.

Credit for statistics

"The second area is credit. The consumer credit market is becoming incredibly competitive and volatile with all sorts of different ways of borrowing and repaying money. An improvement of half-a-percent in a market of hundred of millions of pounds can make banks a lot of money. Mathematical sophistication can make a substantial difference in revenue.

"That's a way applications have changed statistics and continue to change statistics, leading to new statistical techniques by presenting new kinds of problems.

"But then there is this immensely important development of the computer totally revolutionising statistics…

"In the old days – and this is still the popular image – your statistician was someone a bit like this old-fashioned clerk adding up a table of numbers, scratching away with a pen. This is why statistics are still regarded as intrinsically boring by many people.

"That's a total misconception nowadays. The computer has revolutionised statistics. The statistician is more like an explorer. You have a mass of data and use very sophisticated tools in the computer to probe for structure or patterns.

"You don't have to know arithmetic – the computer will do it for you. What you do need to know is how to pick your tool and interpret the results. The computer is leading to entirely radical developments. Things you could do but would take three months you can now do in a split second.

"Previously you would have to be damned sure you wanted to do it before you committing yourself to three months. Now you might do it ten times in slightly different ways because it doesn't take any time at all. That's not necessarily an unqualifiedly good thing because it means you're not thinking so carefully about what you're doing. But obviously it has the potential for good because you can look at data in many different ways.

"So, that's old techniques done in different ways. But there are also entirely new techniques not imagined before the computer: for example, building very sophisticated models in the area of econometrics. You have huge systems of linear equations which you couldn't manipulate without a computer.

"On a more mundane level, there are smaller models for smaller datasets without explicit solutions. To find solutions you have to use iterative methods. The computer allows this – a result in no time at all. And new classes of models can now be developed. There was no point before because you couldn't estimate the parameters.

"Other techniques, often called computer intensive methods, take a sample of data and reuse it in various ways, taking sub-samples. This enables you to say not only *this is a good estimate* but also how accurate it is, without making gross assumptions about distributions and so on – assumptions you can't test. You might look at a set of data a thousand times from different angles. Before the computer this was just inconceivable.

Technology unlimited

"With any scientific problem you start with an abstraction filter. You ask what's important and build a model to represent the relationships between important factors. The computer allows much more elaborate models.

"Nowadays you don't have to take just the three most important things. You can take 100 things. With a small dataset you couldn't estimate the relationship between 100 things very

Does it work?

*Says **Hand**: "Most of my work in the finance area is on credit scoring – trying to produce more complex models. If you look at the history of consumer credit, all the models the banks use at present are simply concerned with predicting default risk. But the banks really aren't interested in that. They are really interested in profitability. Someone might default on their loan but not until they've paid back enough to be profitable. Or someone might be a very good risk but if they pay off their credit card every month they are not profitable. These are more complex issues and I have researchers working on different aspects.*

"Banks are very alert to new statistical methods. Most have explored neural networks. Before that they looked at expert systems. They are looking at genetic algorithms. With the single market and globalisation, they can't afford to lag behind.

"But they're very much driven by practical application not theoretical niceties. The key question is: does it work – will it give us a competitive edge? But it's a very volatile industry, so your competitive edge doesn't last very long. Like the red queen, you have to keep running to stay where you are. That's good: it provides yet more impetus for scientific development.

"I'm also involved with various pharmaceutical companies – again, primarily developing new models or applying existing models in new ways.

"This is mainly in clinical trials. One example is synergy. Nowadays you hardly ever find a single drug prescribed – it's almost always a combination. This might be one drug to impact the symptoms of the disease and another to fight the side-effects of the first drugs. Sometimes two drugs together work better than an individual one. I have been looking at some very interesting theoretical questions there."

accurately. But you can with the huge datasets now possible.

"We can consider much more subtle relationships. For instance, previously we often only had the mathematical and computation ability to assume that things were independent. Now we can relax that restriction.

"A good example is the colossal amount of work over the last 10 years on repeated measures data: these are data in which each subject is measured more than once, so you might have 30 observations on each of 100 people, for example. Obviously there are relationships between repeated observations and you need to model the correlations.

"A huge range of techniques has been developed to introduce the extra sophistication for relaxing restrictions previously ignored because you couldn't cope with them."

Surely, I ask, there is a limit to such developments?

Says **Hand**: "Statistics have been described not as science but technology – and I think technology is unlimited.

"We're always encountering new problems to which statistics can be applied, and there's no limit to refining a model of something. Take a ball thrown through the air. You can model it as a simple parabola and then you can take account of air resistance. And you can go on like that for ever.

"We're also driven by this extraordinary development of the computer. No end in sight there. So these things *may* end – it would be foolish to say *never* – but not in my lifetime or my children's."

A question that worries me is that although commercial interests may be benefiting from all this sophistication, what about the man and woman in the street?

Hand: "I think the banking developments are two way. Say someone wants credit, the bank applies one of these statistical techniques to their application and decides it's unwise to lend them money. This isn't *just* a question of the bank not making a profit. It's protecting the applicant as well.

"There are many such areas where statistical applications have a direct impact on individual well-being.

"I think that, in general, statistics like science are morally neutral. You use them how you use them."

I ask if there isn't a danger of data overload? Can we ever make sense of them all?

Hand: "Maybe it's like never having too much money – you can never have too much data. Whether these data can be used constructively is another question.

"You do get situations where we have collected all these data and missed some crucial item so we can't answer a vital question. I've seen someone collect data for five years for a PhD and then seek statistical advice on how to analyse them, only to be told, 'You can't – a pity you didn't see a statistician five years ago!'

"But data don't cost a lot to keep so I don't see that having too much should ever be a problem."

Faith in statistics

We discuss major problems that advanced statistical techniques have identified, and the action taken. The insupportable volume of road traffic, for example: surely the data on this are overwhelming but are we taking heed?

Says the **Professor**: "I'm sure that within the next decade we'll see the price of petrol go through the roof as the squeeze is put on car ownership. Car manufacturers are

A world of astronomical datasets

Says Hand: "One important impact of the way Europe and the world are developing is fusion of datasets, for example from different national statistical bodies that collect things in slightly different ways. This is opening new areas of statistical investigation, like metadata. There are issues of how you can manipulate metadata, so, for example, one can take unemployment statistics collected in different EU countries and produce a sensible well-defined overall figure.

"If I were advising someone on an area for future for research I think a good bet would be large datasets. You now have astronomical datasets – not just thousands, millions, tens of millions, but hundreds of millions of records. This needs new solutions, new technology."

developing technological solutions to this – radar systems that keep you a fixed distance from the car in front. Clearly such solutions are not an ultimate answer but might ease things and allow more cars on the road.

"But I think the point is that the problems have now been recognised. I'm sure they will come up with a viable solution. Society *does* adjust to pressure – I don't think we'll suddenly hit a brick wall. When things become too intolerable in one direction a different way is found. I guess I am not ultimately pessimistic about this.

"Statistics are crucial because the models enable you to see that certain solutions are not viable. There are classic examples of modelling the traffic flow on motorways to explain why as you drive along, the cars slow down, then stop, then gradually start again – and there's no apparent reason. It's just the way motorway traffic behaves.

'Eurostat's head screwed on'

*"In some ways when it comes to R&D", observes **Professor Hand**, "Eurostat are too far ahead of the game. I think this was the case with the statistical expert systems, DOSES. But they are doing the right things and providing resources for research.*

"One problem is the difficult but understandable one of trying to achieve two objectives. One is to solve problems they have identified quite rightly as confronting us in the not-too-distant future – further unification, merging datasets, sheer size of datasets, different definitions, and so on.

"The other is quite proper concern with stimulating collaboration across international boundaries. Such projects are not always without their problems of working smoothly. A lot of resources are consumed just travelling around Europe. It's not always clear who owns the project so there can be lack of commitment which can hinder the success of the technical content.

"I don't think the structures set up for such research projects have always been ideal. There are no easy solutions but the difficulties need to be overcome.

"But they are identifying problems that must be solved before too many years pass. I think Eurostat basically have their heads screwed on – are going in the right direction. Nobody can get everything right; I think they are doing pretty well."

"So we understand that we can now get round that particular problem if you keep all traffic travelling at a certain speed. We needed a mathematical/statistical model to find that solution. You can't do it without the statistics!

"The trick is identifying a problem that will become serious at first glimmer or hint that it's happening. By the time you have a major epidemic, you *can* recognise it and in a sense are doomed to failure."

What, with his statistician's eye, does he see as other looming global problems?

"The world is shrinking: more and more communication, more and more travel, easy travel. I think problems may arise from this. We shall cope by developing technological solutions but any solution always has a side-effect. Our mathematical and statistical models will show how all these things interrelate and enable us to detect problems ahead and cater for them.

"You need teams of statisticians and domain specialist working together. There's definite danger of compartmentalisation at the moment.

"At the start of any big change we are looking for small effects that are going to grow; to do this you have to have a big sample size. A place like Eurostat that spreads its web throughout Europe is ideally situated.

"I have faith in the power of statistics!"

What's in a name?

*"There are quite a lot of people working in statistical R&D but many don't call themselves statisticians", says **Professor Hand.** "A lot of new methodological developments in this area have been by computer scientists, artificial intelligence researchers and people in other disciplines. Expert systems could have been developed by statisticians but for one reason or another were not.*

"In neural networks there's a lot of overlap. People have recognised the value of statistical ideas, but when the work started statisticians weren't involved. Statisticians have evolved similar things. Where they went wrong was not calling them by a name such as neural networks which was bound to attract attention. The trick seems to be to come up with a name like expert systems or neural networks and all the media take an interest. If all you come up with is generalised additive models people are not so excited."

JOHN WRIGHT *also talked to* PROFESSOR ANTONY UNWIN, *a specialist in interactive graphics at the University of Augsburg in Germany. He found another academic statistician heavily influenced by the need to solve real problems. He believes academics should do more to keep the subject down to earth. But, above all, his philosophy is that...*

Data analysis is fun

"A main theme of my work is that data analysis is fun", says **Antony Unwin.**

He explains: "It was hard to convince people of that in the days of mainframe computing when it was difficult to get data into the computer for analysing and when results were only available, badly printed, on huge sheets of green and white paper. What do primary schoolchildren draw on these days now their fathers don't bring home piles of old computer paper?

"Nowadays, much data is automatically stored on computer and it's easy to display results in an attractive and understandable way. Interactive graphics software is especially good for this and very attractive to work with.

Interactive graphics have changed how statisticians explore data. They will also change how statistical concepts are taught. Through live visual displays interactive graphics offer insights which are impossible to achieve with traditional teaching tools. This obviously affects the teaching of service statistics courses, which can be made more attractive and accessible, but it also affects advanced courses for those specialising in statistics.
Antony Unwin.

"Everyone can grasp the ideas in well-designed graphics. Interacting directly with the pictures to get further information without complicated command sequences makes everything much more accessible.

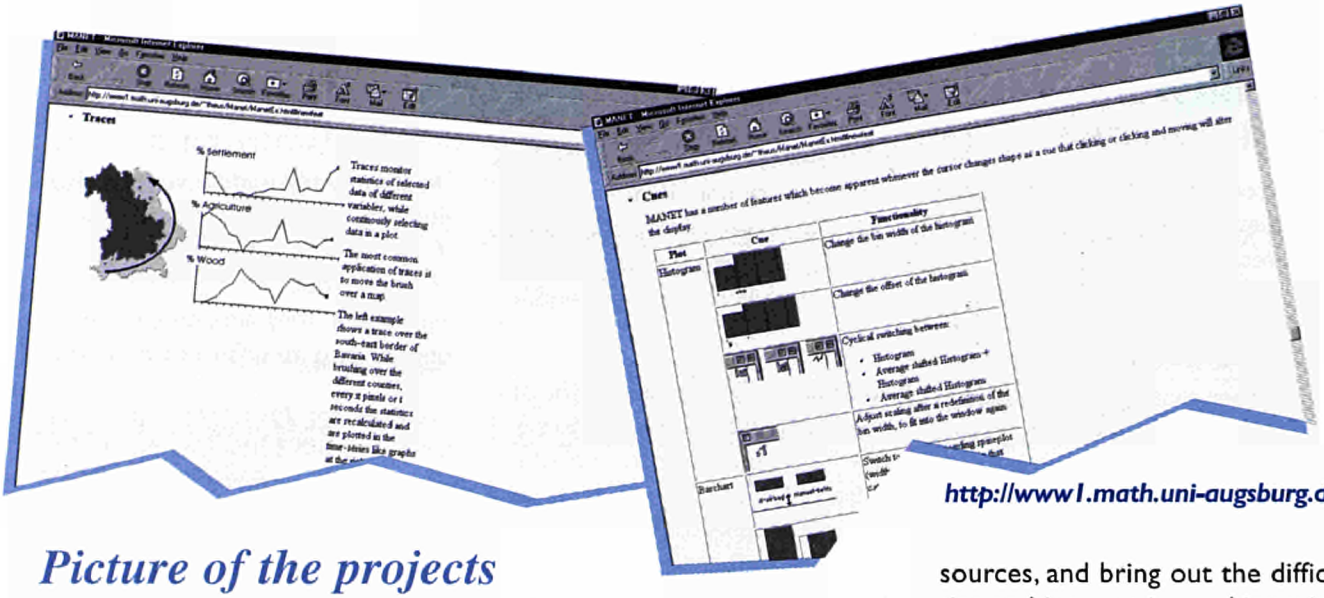
"Data analysis means not only analysing the data but incorporating all sorts of other knowledge and information into the interpretation of the results. This requires the involvement of people who know the background well and interactive graphics are an excellent intermediary."

He warms to his theme: "Like any other enthusiastic specialist I tend to see my subject in everything. There are fascinating examples in medicine, in social statistics – even in music and sport. Anyone opening a newspaper needs to understand statistics and to be able to cast a critical eye on the usually appalling statistical graphics that newspapers seem to offer as decoration rather than as information.

"Last term my department ran a seminar on sports statistics and looked at soccer results, ratings of ice-skaters, volley-ball tactics and data from several other sports. My group has looked at Eurovision Song Contest voting, geographic patterns of dialect in Schwaben, voting in Irish referenda, speech recognition data, commodity prices, risk management, lottery results..."

***Antony Unwin** is Professor of Computer-Oriented Statistics and Data Analysis at the University of Augsburg. This is a new chair started by initial funding from the Volkswagen Foundation. He is interested in exploratory data analysis, interactive graphics, spatial statistics and the developments of statistical software. All his work is heavily influenced by real applications and he has extensive experience of cooperating with medical researchers, scientists, industry and government. He is currently chairman elect of the European Section of the International Association for Statistical Computing. He participated in the first research programme supported by Eurostat, DOSES and in the LIKELY project, and is a reviewer for the second programme, DOSIS.*

Although he has spent most of his working life in academia, including 14 years at Trinity College, Dublin, he has also worked in industry. He says: "This experience was valuable in giving me a much greater appreciation of the difficulties faced in dealing with real problems."



http://www1.math.uni-augsburg.de/ ~ unwin/

Picture of the projects

Professor Unwin's current projects in the Department of Computational Statistics and Data Analysis at Augsburg include MANET and TURNER.

***Manet** is software for interactive statistical graphics running on Macintosh computers. It provides the standard interactive graphics tools extended to deal with datasets with missing values and additionally offers many new interactive features. All graphics are fully linked and may be interacted with directly. MANET follows the Macintosh conventions and is consistent with other Mac packages. It is an exploratory tool to be used in conjunction with other, more traditional software.*

***Turner:** Interactive contingency tables provide the user with the facility of easily switching between low-dimensional views of the multivariate data. Using raw data as well as local and global summands of the power-divergence-statistics, they identify unusual cells that have a strong impact on the resulting statistics. Interpretation of significant effects is straightforward by combining categories and collapsing tables. TURNER, running on Mackintosh computers, offers all those interactive features as well as easy-to-use log-linear modelling facilities.*

Off course

Professor Unwin continues:
"There are many statistics courses but few mention official statistics. There are many reasons for this. Statistics' courses are commonly taught by mathematical statisticians who concentrate on mathematical theory. Other courses are taught for scientists: biologists, geologists, physicists etc Even courses taught for social scientists often concentrate on how data sets may be analysed but not on how they are collected.

"I occasionally get the feeling that overly mathematical arguments and formalisations are gratuitously included in otherwise reasonable texts to give them some kind of intellectual cachet. The real problems lie often not in the mathematical theory but in the down-to-earth difficulties of data analysis.

"David Moore's book *Concepts and controversies in statistics* is an excellent example of how statistics should be presented at an introductory level. His title genuinely reflects his book (unusual in a world of academic texts, where 'applied statistics' or 'applications of statistics' in the title may bear little relation to the contents). The examples are real ones, from a wide range of sources, and bring out the difficulties and interest in working with data and statistics.

"The best description of the problem with statistical education is by the famous statistician George Box. He suggested that teaching statistics in the classroom was like teaching swimming in the classroom. If you never had a chance to practise in the pool you would only be good for teaching more people in the classroom.

"I would extend the analogy further. Swimming pools are calm, of known and even depth and the water is clean and warm. With a bit of practice you could swim blindfolded in a pool. In terms of statistics teaching it would be like working with data sets but only nice, tidy ones: those that had been cleaned up, were small and well-defined and enabled you to test very sophisticated mathematical models.

"But real statistical analysis isn't like that...

"You have to organise the data, check them, edit, reformat, adjust... Maybe you have to collect more data. A lot of work has to be done before you get anywhere near applying the mathematical methods which take up most of the space in the textbooks."

Academics, industry & official statistics

Unwin continues: "Statisticians in universities and research groups are interested in analysing real problems – but usually they want new and, ideally, technical problems which can be well defined and solved, rather than methodological problems, which are vague, hard to specify and possibly impossible to 'solve'.

"When computing resources were slow and limited, technical problems of this kind were very apparent. With the speed and performance of modern computers there is more emphasis on what place techniques have in an overall strategy of data analysis. Waiting two days for the results of a regression analysis or even having to wait ten minutes is quite different from having instantaneous response. Many more analyses can be carried out. In particular, more thought can be given to graphical investigation of data and models, to alternative models and to sensitivity analyses.

"Official statisticians are concerned very much with the qualitative nature of their data – what is referred to as metadata. They want to know why the data were collected, the definitions used, when the data were collected and how... None of this matters in the theory of a mathematical statistician.

"Metadata structures are most valuable when data sets of similar kinds are collected – when there is much repetition and when standards are essential. All these factors loom large in official statistics and in industry but less so in scientific research.

"Eurostat kindly gave a student of mine access to data on part-time work in the European Union. On the face of it, there were equivalent data on a number of factors for each country. In practice, the definitions of part-time work vary enormously across the Union and any attempt at comparative analysis can be made only with the greatest of care and with a large number of qualifying statements.

Two extremes

"At one extreme academics would like research programmes to support their research in a general way, although most are realistic enough to recognise that that isn't on. At the other extreme, funders of research would like particular problems solved. A perfect meeting of minds is rare.

"What we do find is that particular problems are not quite what they seem and that it can be better to do something completely different, ignoring the original problem. We may also find that the stimulation of a specific problem suggests ideas that are of much more general application and of wide research interest.

"In the muddy middle we can also find projects where the research does not match the problem and, as both groups shift ground to find agreement, they sink deeper and deeper into an unsuccessful mess."

He goes on: "Most of the projects have involved software implementations of the research ideas. Software development still seems to be a question of promising everything and delivering little. With standards changing so quickly it is hard enough, but maintaining progress in an international project with partners of equal status is even tougher.

"There are two conditions of European research funding which are both eminently sensible and yet awkward. The first is that researchers from different countries must cooperate in a project for it to be funded. This should encourage cooperation between countries (which it does) but also leads to large travel expenses and project management complications.

"The second is the involvement of industrial partners. In both the original DOSES and now DOSIS, Eurostat asked for proposals for research in areas of official statistics. While some of this work is highly relevant to broad areas of statistics, some is specific. And there are many areas of statistics not covered and which would be of prime interest for possible industrial partners.

"Consequently the partners outside universities have tended to be national statistical offices. This has not been entirely in line with the aims of the framework research programmes. But industrial partners are reluctant to participate and invest in projects where there is a restricted market for the resulting products."

Analytic methods have dominated statistical teaching. Interactive graphical methods are now available which complement and enhance the analytic approach. Interactive graphics are accessible, attractive and effective. They should take a central role in teaching statistics in service courses to make statistics comprehensible with graphical explanations and graphical explorations.

The proposal is not to discard analytic methods but to place them in a context of understanding. As they have always been, analytic methods are an important counterbalance to jumping to conclusions. **Antony Unwin.**

'The world is multivariate not univariate and statistical teaching must reflect that. Interactive statistical graphics is the key.'

Antony Unwin

Official statistics 'neglected by academics'

Unwin adds: "I have mixed feelings about this, because research into problems associated with official statistics has been neglected by academics.

"As with many academic subjects, the most kudos can be gained in exact manipulations of sophisticated formalisms, whereas real problems demand more practical compromise solutions.

"Official statistics are crucial to the flow of information in society. They are one of those subterranean pillars that are relied on but never visible.

"It will be interesting to see if the new journal *Research in official statistics*, started by Eurostat, will be able to affect this. Ideally, it should attract, on the one hand, outlines of practical applications that stimulate more relevant research; and, on the other, descriptions of new research that could be applied in these fields.

"Anything that brings users and researchers closer together must be viewed optimistically.

"One of the interesting threads in European funding is the supporting of international workshops", **Unwin** continues. Eurostat started a series on neural networks a few years ago when that was a hot topic and last year also supported a workshop here in Augsburg on *Strategies for data analysis* (see *panel*)."

He concludes: "Planning research programmes and allocating funding is a difficult and thankless task. There are very few major successes in any field, it is extremely difficult to identify the reasons for success – and even if they have been recognised it is difficult to recreate them to generate more successes."

'Opportunity for Europeans'

A workshop on Strategies for data analysis *was held at Augsburg University last October. Says* ***Unwin****: "Thanks to the support of Eurostat it was possible to invite many of the leading researchers in statistical software. The key element was the joint analysis of datasets provided by the participants.*

"The datasets were not 'classical' test cases, but datasets for which the essential background information was readily available. Dataset owners were present to describe the analysis needs and the context. So not only did we have talks on new research developments in this exciting area, we also had the chance to see many different software packages in action on the same real datasets in the hands of experts."

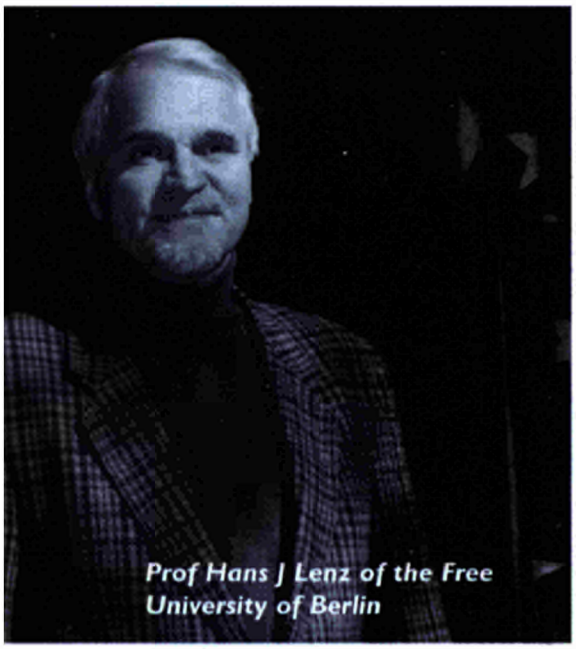
The meeting was opened by ***Professor David Hand*** (see article on page 20) *who gave a keynote paper on the importance of understanding the substantive questions to be asked of datasets.*

Dataset discussions highlighted several key issues in data analysis and in the use and assessment of software:

- *Datasets are analysed informally under the guidance of few generally accepted principles. There is always more data preparation, checking, editing and reorganisation than anyone expects, especially with complex, real datasets.*
- *Software must be flexible, offering manipulation tools, good graphics and excellent connections with other packages.*
- *Analytic and graphic methods complement each other well in the analysis of datasets but not well within packages.*
- *Exploratory data analysis is a substantially different approach to confirmatory modelling and the packages reflect this.*
- *There is a continuum of software needs and software users. No one current package can cover the whole range.*
- *There is a learning curve for any software and there is a tendency for people to stay with packages they know rather than to invest more time in learning another. This has important implications for the kind of analysis that may be carried out and the strategies people adopt.*
- *There is an opportunity for Europeans to develop the next generation of statistical software. Many of the innovative ideas are European and Europeans are not as heavily encumbered with the need to maintain compatibility with previous generations of software.*

In this article PROFESSOR HANS J LENZ, Free University of Berlin, comments on federated statistical databases and improving the interoperability of global and local database systems.

Sharing data globally



Prof Hans J Lenz of the Free University of Berlin

A federated statistical database system offers a solution for the collaboration of heterogeneous, autonomous and distributed database systems. It can be viewed as a compromise between two extreme solutions: total standardisation and absolute autonomy of the local database systems. The heterogeneity is caused by conflicts of hard and software and data models of the various sites. It includes, of course, the hard problem of semantic conflicts.

Consider the EU national statistical institutes and Eurostat: to produce and publish summarised and integrated European statistics, each national or local site must filter out the needed data, transform the filtered data into a component schema and transmit the transformed data to the supra-national or global site.

Typically the local database systems (DBSs) are geographically distributed, autonomously operated and heterogenous with respect to software and hardware and the underlying data model. Moreover, there exist semantic conflicts. For instance, definitions of attributes and domains of nomenclatures, code lists and classifications may vary between NSIs.

The same kind of conflict happens at a lower level – within a nation. Now consider the DBS of an NSI as a global site and take the DBSs of companies that are reporting periodically to the NSI as local sites…

Assume that the local sites and the global site are connected by a communication system based on dedicated or dial-up lines or the Internet. Assume further that the necessary level of network security and data protection is guaranteed.

A federated database system (FDBS) is a collection of cooperating but autonomous heterogeneous local database systems that allow partial, controlled sharing of data on a global level.

Such a federation can be achieved by a 'stepwise' procedure. The divergent local schemas are transformed into a common data model (*component schema*) which all sites agree. Because only a proper subset of data is needed from each local site on the global level, filtering is used to provide access to the shared data and its encapsulated operations. The corresponding schema is called an *export schema* and differs from local site to local site. Finally, the multiple export schemas are integrated in a federated schema.

A challenge

It's a challenge to explore and solve the various kinds of semantic heterogeneity by properly defined primitive operators or reference tables applied as part of the transformation and integration step. There exist hard conflicts that can't be solved by looking up tables or invoking procedure.

Think of a structured attribute like the classification of branches or consider a survey with its specific sampling frame. If the structure of the classification of the local sites and the global site don't match or the sampling frames are not comparable, it's difficult to find mappings to derive the appropriate summary data. In all these cases standardisation or harmonisation of definitions, methods of data collection and models of estimation of aggregates is inevitable.

The research under progress is a collaboration between the Department of Statistics and Econometrics, Free University of Berlin (Professor Hans-J Lenz) and the US Lawrence Berkeley National Laboratory, Data Management Group (Dr Arie Shoshani). Prof Lenz is mainly interested in the intersection of statistics and computer science. Fields of his current research include artificial intelligence, statistical databases and metadata driven data warehouses, and computational statistics.

A statistical eye on R&D and innovation

by Barbara Jakob

R&D and innovation have a crucial influence on economic growth, competitiveness and the postion of economies in the world market. Data on innovation and R&D – their state, development, and composition – are of key interest to decision-makers at all levels.

The statistics of Eurostat unit A4 (R&D, methods, data-analysis) cover these two main fields of R&D and innovation. Says **Maurits Pino**, project manager: "Firstly, all R&D statistics provide information on inputs – money spent and people devoted to R&D. They record the performance of R&D. As an important indicator of the trend, we also obtain information on budget provision by central government for R&D – and form a view on the level of funding planned by government. It is also valuable to identify those areas of research most targeted for funding by Member States and the Commission. All this serves to coordinate the R&D policies of all the different parties involved.

Focus on the regions

"A special aspect of our statistics is the ability to give a regional breakdown. This is a big advantage."

But focusing on R&D inputs is not enough to shed light on R&D and the innovation process. The result of research and development often is a patent. Patents' data enables conclusions on innovative potential. So patents were chosen as one indicator of R&D output.

Data from the European Patent Office are used. "In providing data on European patent applications at regional level we are able to give a first impression of inventive activities and potential within European regions", explains **Sabine Gagel** who is working on this project.

Maurits Pino again: "As well as breaking R&D expenditure and personnel data down to regions, we also included patent data. So we are now able to provide important indicators for R&D input – personnel and expenditure – as well as output – patents – each at national and regional level.

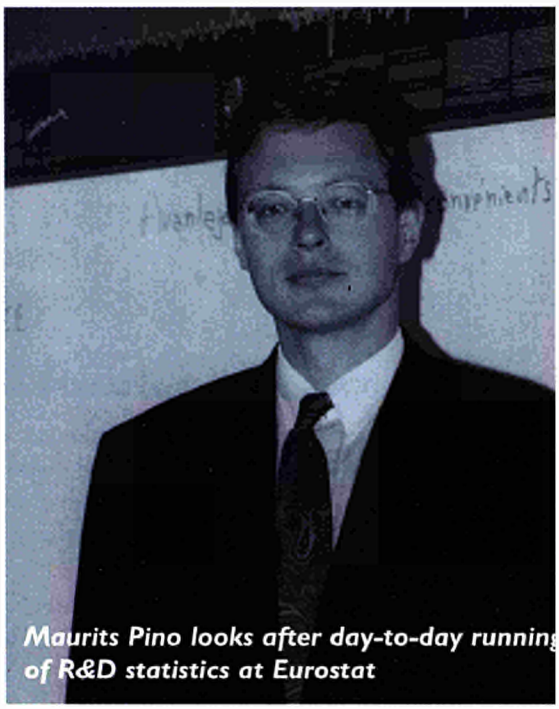
"And then we are in a process of making an inventory of the methods of assessing regional R&D performance. We asked Member States actually how they compile the data and are analysing the results to harmonize these statistics further."

Achievements

In the last few years there have been some major developments. **Maurits Pino**: "In the early 1990s there were only data available on government funding of R&D. Recently, we made an analysis of how Member States report on government appropriations. The result: still discrepancies between methods used. Further harmonization is needed.

"Under my predecessor Werner Grunewald, Eurostat assembled regional data on personnel and expenditure. In 1996 we first dealt with patent data.

"We launched a stock survey on human resources in science and technology, a wider concept than R&D personnel. This survey has had interesting results. We developed this survey on the basis of a joint Eurostat/OECD manual. When we started our survey, the OECD secretariat initiated its own for member states not in the EU."



Maurits Pino looks after day-to-day running of R&D statistics at Eurostat

Innovation – the big issue

Innovation is one topic firmly on the agenda since it was recognised that the financial or physical state of an enterprise is not the only indicator of its chances of success. Many other factors come into play: the research it conducts, staff expertise etc – and intangible activities and investment, of which innovation is part.

In 1992/93 Eurostat launched the first Community Innovation Survey (CIS) as a joint project with the European Innovation Monitoring System (EIMS) of the Commission's DG XIII (Telecommunications, Information Market and Exploitation of Research). Aim was to fill the gap in information on resources required for innovation,

Mikael Åkerblom is responsible for innovation statistics, especially the second wave of the CIS

the way enterprises acquire, develop and transfer technology, and the role innovation plays in development of enterprises and growth of industries.

Using a harmonized EU questionnaire based on OECD methodology, this survey resulted in some 40,000 responses from enterprises. Data are comparable for ten of the 13 participating countries (Greece, Portugal and the UK do not strictly compare because of partial information or too-low response).

The first wave is now finished. Eurostat has just issued a CD-ROM with detailed survey information.

A big pilot

"The first wave was more or less a big pilot survey", says **Mikael Åkerblom**, who is responsible for the project. "Experiences were mixed but generally good in the way the whole concept seems to work. But there's a lot to do in terms of improving harmonization of the data. We shall try to realise the lessons learnt in the second wave of the CIS to be launched in due course.

"We had problems of international comparability due to several contractors' modifying the questionnaire, and sampling methods were not identical", **Åkerblom** explains.

New challenges ahead

Technological change and knowledge are more and more becoming determining factors of productivity and competitiveness. Such change is calling for new statistical approaches. "As for these new emerging concepts such as knowledge-based industry and technology transfer, we are working on the data we have and trying to improve the way we use them. And then we are developing these issues further in cooperation with the OECD secretariat", **Maurits Pino** says.

Mikael Åkerblom explains how these new concepts are taken into account in the framework of the CIS. "The second CIS will tackle these issues, at least partly, especially networking – how much various institutes are cooperating on innovation – and also, to a certain extent, technology transfer.

"The connection with knowledge is very interesting. There is a possibility of linking innovation surveys with information on the knowledge base of enterprises. That could be done in some Member States, though it would be difficult at Community level. Maybe in the long run?"

Core of the debate

Eurostat is also contributing to a large publication by DG XII (Science, Research and Development) – *the European report on science and technology indicators*. In the second edition, due at end of the year, Eurostat is responsible for the comprehensive statistical annex. This gives the most important science and technology indicators which are put into the context of more general economic information about competitiveness.

An aim of the report is to compare the performance of EU Member States in science and technology with that of other leading countries. It is not confined to describing EU or OECD countries. In providing comparative data on some 60 countries, including, for instance, Mediterranean as well as some of the largest developing Latin American and East Asian states, it goes much wider. It thus provides an insight into the scale and effectiveness of efforts in science and technology.

Current developments in society and its economies require statistics to react quickly and effectively to satisfy the need for up-to-date and relevant information. Eurostat's R&D unit aims to provide such statistics.

For more information about the results of research projects and developments in R&D statistics see Eurostat's new bi-annual journal Research in official statistics. *This covers information about the REDIS project (Research, development and innovation statistics) and the statistical research programmes DOSIS and SUPCOM.*

SOME KEY DIFFERENCES

Among Member States, there is wide divergence in terms of socio-economic objectives of government-funded R&D: the British and French devote 40 and 30% respectively of their publicly-funded R&D to defence; the Danish and Dutch allocate most to environment (4%), while the Commission focusses on R&D for energy and industrial production and technology (together over 55% of the total).

R&D tends to be fairly concentrated. In regions near the capital (Munich, Milan) there are R&D intensities far in excess of the national average.

When breaking down the number of patents by international patent classification, one notes that among the large Member States the degree of specialisation is low. The smaller ones tend to be relatively specialised: Sweden and Luxembourg in performing operations and transport, Portugal and Ireland in human necessities and Belgium in chemistry and metallurgy.

For our latest profile of a national statistical office Sigma's JOHN WRIGHT *went to Rome to meet* PROFESSOR ALBERTO ZULIANI, *President of the Instituto Nazionale di Statistica (ISTAT). He met a man determined on and with a novel way of...*

Putting statistics into the hands of the people

He has a passion for putting statistics into the hands of the Italian people. And he's done so quite literally. Ten million times.

Ten million of these L.500 coins featuring the ISTAT building were minted for the 70th anniversary

In his grand office in ISTAT's rather grand headquarters in the narrow Via Cesare Balbo in the heart of Rome, the **Professor** *explains:*

"Last year ISTAT celebrated its 70th anniversary. We arranged for the issue of a L.500 coin with the ISTAT building on one side. Ten million were minted. Why? So citizens could have an image of ISTAT in their hands.

"There was also an ISTAT postage stamp – with a first-day cover. And a telephone card. And a 30-second advertisement on TV after the evening news.

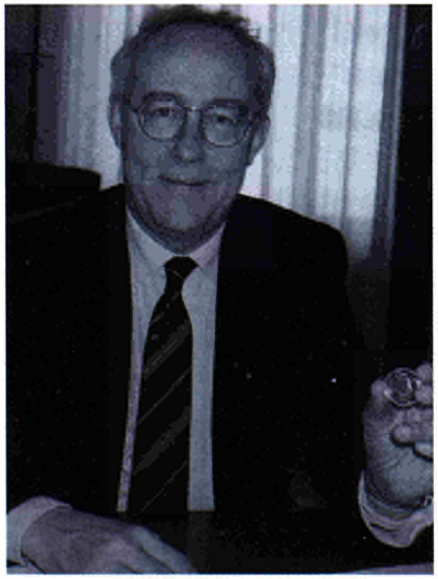
"The slogan was '*For 70 years ISTAT has been reflecting on Italy and helping Italy to reflect on itself*'."

And there's more. ISTAT have produced a book, Profile of Italy, *which, in NSI terms, is a 'best-seller' – 10,000 copies sold in less than two months.*

"I have invested a great deal in reaching out to the citizen during my four years as President of ISTAT", he says.

So just what's been going on at ISTAT? And how does a distinguished Professor of Statistics come to be talking like a high-powered marketing man? Let's take a deep breath and take it from the top, as they say...

Alberto Zuliani explains: "The law decrees that the President of ISTAT is a university professor. I am Professor of Statistics at the University of Rome – I've been on the teaching staff since 1964. I came to ISTAT in June 1993; the appointment is for four years.

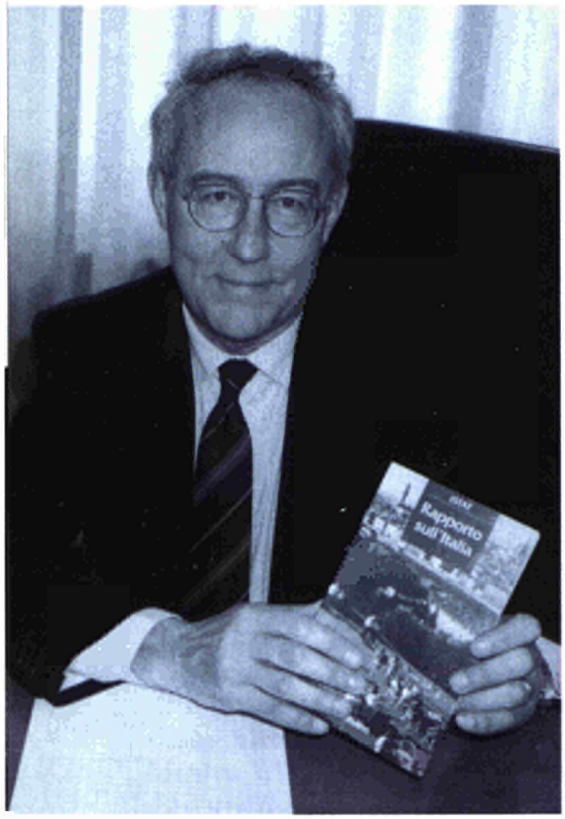


"I had some experience of public administration: I was chief of the department monitoring government policy for 20 months in 1981-82 during the government of Mr Spadolini. Then, as now, I was on loan from the university.

"I was chosen by the Council of Ministers – the Cabinet. I had been President of the Italian Statistical Society and was President of the Guarantee Commission for

▲ ***The ISTAT building in the heart of Rome. The statue is Seshat, Egyptian goddess of numbers. Ancient Egyptians and Romans always faced problems of sharing and dividing. So statistics were part of their lives***

◀ ***Professor Zuliani with one of the coins***



Professor Zuliani with his 'best-seller' Profile of Italy – 10,000 copies were sold in a few weeks

Statistical Information, so I was quite well known in the field."

What did he think when asked to be President?

"Pleased. I was familiar with ISTAT so thought I could do quite well."

What were the big things he wanted to achieve at that time?

"I had some ideas but learned more about ISTAT for some months and then made a plan."

Tell me about the plan?

Big changes

"The main problem was the institute had just changed its main objective from collecting data to being a research institute. Big changes were afoot in the way we operated and in personnel.

"Another challenge was the institute wasn't recognised as playing a key role in the life of the country, but had the possibility and skills to do so. So the challenge was to invest in communication with users, citizens and enterprises and also with policy-makers.

"Like many statistical institutes, we were oriented more towards supporting policy-makers at government level. The aim was to make statistics more relevant to ordinary people and their needs.

"A lot has been done in the past three years towards these aims.

"Now there is much more analysis of the data. One of the most significant changes was drafting an annual report on the situation of the country based on official data. We have just published the fourth. It is a very in-depth analysis of what has happened in the year in the economic, social, environmental and demographic fields."

Such annual report was the basis for the 'best-seller', which he is clearly very proud of…

"Last year for the first time the annual report was summarised in a short book for sale in bookshops and libraries throughout country. We worked with a publishing company, *il Mulino*, and they have sold 10,000 copies – a very substantial number for the Italian market. It costs L.15,000, less than ten dollars. It was presented at various book fairs – Rome, Milan, Turin… – and publicised heavily by the company.

"We also publish some other books with the same company – for example, statistics on the elderly – and they have asked us to publish many others.

'Books more democratic than the Internet'

"Books like this are the way to reach the citizen. Of course, we use the Internet and we have statistical information centres located in each region of Italy. But books are particularly important because the Internet, which may be *very* important in the future, still only reaches certain segments of the population. It is not spread on a democratic basis – doesn't reach everybody. It goes mainly to those who are more educated, those in work, men rather than women, the young rather than the elderly. This is true also for books, but not as much. A book may be less diffuse but as a tool it is

ZULIANI ON 'IMAGE' ...

Is he well known in Italy?

"You'll need a survey", he laughs.

Does he appear on TV?

"Yes. I don't look for it so much – although it could be done much more. Instead, I always try to separate the personal aspect from the institutional one. I'm interested that ISTAT comes over rather than me as a person."

What do outsiders think of all the publicity for ISTAT?

"There's a lot of appreciation and a lot is asked of ISTAT. I believe that in the political world the image of ISTAT has grown considerably."

Where did his passion for communication come from?

"Comes from the reality that even great capacities and capabilities if they don't emerge will not be recognised. When I joined ISTAT I realised that it had a lot of capabilities but there was also a lot of criticism. There was no recognition of the importance of information – the importance of statistical independence – so I thought it was important to invest in this."

more democratic. However, we have also other tools, for example an information page on TV.

"The aim of this book is not to explain what ISTAT does. It is an analysis of the situation of the country from a statistical point-of-view. It never gives recipes, advice. The boundary of ISTAT's research, which is always as neutral and objective as possible, is to explain the situation on the basis of data.

"We also use the media a lot, especially television."

And then he talks about that L.500 coin...the stamp...the TV ad. Why is he so passionate about reaching the citizens?

"It is so important because without statistics there is no democracy – although, of course, there *may* be statistics without democracy.

"This has been the most important aspect of my term in office; I have invested very heavily in this issue. Although there were some early doubts, we have already had some returns from this approach: the credibility of the institute has increased considerably. Just one example: a well-known car advertisement refers to the institute's data. And, contrary to what is happening in other countries, the government has allocated greater financial resources to the institute than in the past."

I ask about the reaction of government to this new open approach. Are they happy with all the publicity; clearly, it's not all 'good news'?

"We don't have problems because ISTAT has always been strongly autonomous from government. From an administrative point-of-view it is subordinate but on the technical side totally independent."

He points at the 'best-seller' again...

"Nobody looks through the draft of a book like this and says 'You shouldn't put this or that in'. "Since I've been here – and since

Statistics for the people: in the datashop, just across the street from the main ISTAT building

Italian governments don't last very long! – I've already worked with four different heads of the council of ministers but no one has ever made an 'improper' request. And, to safeguard this independence, for three years we've published, at the beginning of the year, a calendar of the press releases and all the data to be released from the institute.

"At 8-30 on a release day there is a meeting with all the press agencies with an embargo to 9 when the data are released to the outside world and to government. Before 10 the figures are on the Internet."

But, I probe, could anybody see key data before 9 in the morning? What if the Prime Minister rang him up the day before and asked as a special favour...

***Zuliani's** reaction to this question makes it clear that such a scenario is too far-fetched for serious consideration!*

A question of independence

Our conversation then diverts to wider issues of statistical independence

"This something ISTAT carries forward at European level. We think there must be greater independence of Eurostat. We are a promoter of this."

Doesn't Eurostat have enough independence?

"It has some autonomy but is a directorate of the Commission. ISTAT is not a directorate of any ministry.

"Eurostat's position is different from the one we have experienced historically in Italy – and that in other countries – where the independence of statistics is much stronger. We feel it would be better if Eurostat were more analogous to national statistical institutes in terms of its independence.

"Also it should foster very strong cooperation among and with the national institutes. For example, in our national statistical system ISTAT has a central role but there are a number of other statistical bodies or entities – within ministries, for example, and also at regional level. The strength of the system is cooperation among all these bodies aimed at using, as much as possible, all the statistical information available at their level but not hitherto used to its full extent.

"We did much during the Italian Presidency to evolve the European system in this direction. And may be

Central Rome: As the statistics show, Italy is a mixture of old and new, affluence and its opposite

it will gain some fruits – may be, I'm not sure. I think many national institutes are beginning to have the same feeling: that there should be more integration, more cooperation, less top-down decisions, more bottom up."

Objectives achieved?

We switch tack. He's approaching the end of his term as President – has he, to a large extent, achieved his objectives?

"I had very ambitious objectives. I feel I have achieved them in a very satisfactory way – up to 80 or 90 per cent – especially reaching out to the citizen. So, some fully reached, some not yet...

"One objective I haven't been able to achieve fully is a closer relationship with enterprises. Enterprises pay for statistics by supplying us with information. They don't realise that all the time information is flowing back to them, although at aggregated level through some other body and not directly from ISTAT.

"I have been trying to improve our relationship with them. This year for the first time there will be a multi-purpose survey on enterprises. We've already had such a survey for households. Main aim is to study flexibility in the use of the workforce.

"But this relationship in Italy is quite difficult – more difficult than in other countries – because there is a huge number of small and medium enterprises and few large enterprises. So it's much more difficult to reach them and have a direct relationship. On the other hand, we have very strong relationships with the associations that represent them."

Are small and medium enterprises reluctant to supply data?

"In comparison with other countries both households and the enterprises have a very high response rate."

But, I ask, aren't Italians well known for being 'casual' in their approach to officialdom?

"Relationships with bureaucracy are always somewhat of a conflict but not really with statistics. ISTAT runs some surveys on customer satisfaction with public services, including ISTAT. We ask them how much this relationship costs – in terms of price but also in time taken to answer our questionnaires. And the cost is not very high – at least that's what they say."

Talking of costs, I ask, what about your return from commercial activities – do you make a lot of money from publications?

"Not so much. Because there is a conflict: as an institute funded more than 80% by the state, we can't ask for money twice for the same thing. So there are those things we provide without charge and other things for which we can charge, and they represent the other 20% of our budget. The revenues are not so much from publications but from service contracts – mostly with public bodies, the postal service, TV etc, and also individuals."

What next?

So, he's coming to the end of his four-year term. Will he be given another four years?

Much laughter around the table. **Zuliani** makes a fingers-crossed gesture that aide Cristiana Conti says is 'very Italian'.

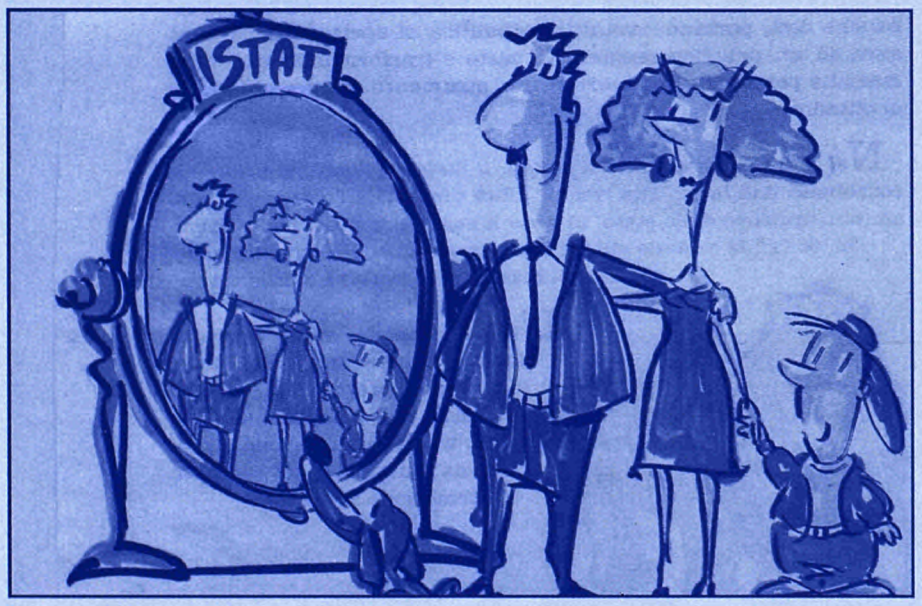
Tactfully, the **Professor** says: "I feel a lot more needs to be done for public statistics whoever is the President

(cont. p 36)

Zuliani on integration of economic, social and environmental statistics

"I am very happy that the IMF has been complimentary about our economic statistics. Now we must take further steps to integrate the data and make fuller use of them. This means working not only on economic statistics but also on environmental and social statistics – all in an integrated way. This is something only an NSI can achieve – only they have so much in-depth information down to individual level.

"I think integration is the future. It is not possible today to speak about economic statistics without considering their impact on the environment. Moreover, here in Italy, with so many small enterprises, the borders between households and small enterprises are very ill-defined. Also the border between profit and non-profit is difficult to detect: the relationship between work and non-work. So to separate the social sphere from the economic one is impossible."



ISTAT slogan: For 70 years ISTAT has been reflecting on Italy and helping Italy to reflect on itself

Zuliani on major changes in Italy in the last 20 years monitored by ISTAT

"One started somewhat later than elsewhere but is coming up very fast – the importance of the services sector. This is a very in-depth and profound change. Second change is in pattern of consumption in terms of level and type. Third is participation of women in the workforce – this started somewhat later but is coming up very quickly.

"On consumption, household costs and foodstuffs are weighted much less than previously in comparison to more evolved types of consumption – health, education, culture.

"And women: Italy was a country where women were not in the workforce. Today the rate is still lower than in other European countries but the increase has been much more rapid and that's one of the reasons why we're seeing fewer children and fewer marriages."

Is this affecting the image of Italians as essentially 'family people'?

"It certainly is. Although the changes may be not as strong as in other countries, for us they are very strong because they started from a very low level. But the family network is still very robust – although nowadays it can be through the telephone or other types of communication – and, as I have said, this impinges significantly on the workplace.

"But families do tend to live closer to each other and marry much later. Often men remain in the family until 30 or 35."

And the 'black economy'?

"We are quite involved in the study of the informal economy and we also assist many countries in transition and developing countries in studying such phenomena. This is very important everywhere. We were among the first to develop methodology for its study in 1988/89. It is now used and applied in many countries and incorporated in the European System of National Accounts."

***Women in ISTAT – now and pre-war. As Professor Zuliani** says, "Italy was a country where women were not in the workforce. Today the rate is still lower than in other European countries but the increase has been much more rapid and that's why we're seeing fewer children and less marriages." Says* Profile of Italy:

"In Italy...the crucial objective of winning recognition of the active role of women both inside and outside the family comes up against a situation which...penalises women first and foremost in the labour market. 1995 data show that young women have major difficulties in taking on the roles that contemporary society asks of them: those of partner, mother and worker." A recent European Community Household Panel survey, reported by Eurostat, said over 40% of Italian women between 25-59 still describe themselves as 'housewives'. In Denmark, in comparison, it's only 4% with an EU average of 33%. A third of these Italian housewives said they were 'dissatisfied' with their situation and 15% were looking for a job. Employment rate of Italian women is under 50% compared with an EU average of 56% and rates as high as 73% in Denmark and 67% in the UK.

for the next four years."

What more?

"There are many challenges but two main ones:

"One is detecting changes in a world changing very quickly – far more rapidly than an institution like ISTAT can change. This is one reason for our multi-purpose survey on enterprises.

"We have changed from an organisation that gathers data to a research institution. This is a transition still being implemented. And it's one made difficult by all the drawbacks of a public organisation. A change of this kind is especially a change in staff; and, while it is easy to put it all down on a piece of paper, the actual change is a complex process. So improvement in quality of staff is certainly one of the challenges ahead."

What, I want to know, is the attitude of the staff to such major changes. Do they take some persuading to go along with him?

"I took my post at a very fortunate moment because it was when a lot of the old staff were leaving. It has been possible to recruit a lot of very interesting staff in a job market which was very stagnant. This has been a big advantage. The last appointment the institution made of a central director – a very important position, highest under the Director-General – was of someone 39. When I entered the institution this was something that was impossible to think of.

"The last board for a research director, also a top position, had six people with an average age of 40 – also impossible ten years ago. Average at that time would have been 55-60.

"In the past four years people in the key positions have changed 60 to 70%."

'A very nice person to work for'

*I then ask **Professor Zuliani** how, outside statistics, he relaxes.*

"I relax with statistics."

Other interests then...

"Sport – I ski in the Italian Alps. And I go to a gym close to ISTAT

Zuliani on how he first became interested in statistics:

"By chance. I studied at the Faculty of Statistics in Rome and had been uncertain about whether to study there or the Faculty of Economics. I learned to appreciate data much later in life – not right away.

"I developed the belief that statistics teaches us to be tolerant because they teach us that in the world things happen in such a random fashion that one cannot necessarily accept one's own certainties as the truth. And to my way of thinking statistics are a kind of poetry, presenting many ideas in a very synthetic way."

An art form even?

"I don't think so", he laughs.

"So I learned to appreciate statistics after my university studies, as I became aware of these aspects. What made me appreciate them even more is that statistical methodology allows in-depth observation of many situations from different points-of-view. Statistics do much to satisfy one's curiosity.

"He or she who is not curious should not study statistics."

to stretch out a couple of days a week.

"My hobby is work. It relaxes me. I like to work and I am very much fulfilled by it. I don't need to relax in other ways, although whenever possible I do spend time with my family."

Obviously he drives himself hard. What does he expect from those who work for him?

"Innovation. Ability to change. Management of human resources. Technical skills But innovation is the most important."

Certain things you won't tolerate?

"If people are lazy: mental laziness, which is exactly the opposite of the ability to change."

Do you think you're an easy person to work for?

"Not for me to say... It depends – some say 'yes', some say 'no'."

Aide Conti chips in: "He is a very nice person to work for. I think you can love him or you can hate him without a middle position."

Most definitely, the middle position is not Alberto Zuliani's style in any respect and Italian statistics are reaping the benefit.

ISTAT in earlier times. *ISTAT was founded in* ***1926****. However a general statistics division was already functioning in* ***1861*** *as part of the Ministry of Agriculture, Industry and Trade with statistical offices in the provincial government. In that year there was the first general census of the population. In the* ***1950s*** *ISTAT sought to provide government with a statistical 'map' of the country's war damage to further reconstruction. In* ***1967*** *the first Italian national accounts appeared. In* ***1983*** *the main databanks were made public. In* ***1996*** *ISTAT celebrated its 70th anniversary. It was a proud moment with the President of the Republic actively involved. And ISTAT made this declaration:* It is important that citizens be aware that data offered for the evaluation of all are, in effect, the data of everyone, collected with honesty, constructed with the greatest possible scientific rigour, and made public as soon as possible

The Regulation on Community Statistics or 'statistical law' came into force in February. Sigma's STEFFEN SCHNEIDER *spoke to man-in-the-know* CHRISTIAN ENGELAGE *about this major development...*

An act of trust

"I am not saying statistics could no longer be produced without this – but..." Christian Engelage, acting Head of Eurostat unit OS-4, speaking at the end of work on the new 'statistical law'.

For about six years Eurostat has worked hard to solve one of the main problems of Community statistics: absence of basic legislation organising the European statistical system at both national and Community level. As a result the Council of the European Union on 17 February adopted what is known as the Community statistical law[1]

There were many reasons. Says **Engelage**: "It was, quite simply, time to come up with some form of basic statistical law – all-encompassing legislation as found in Member States, which would define the basic conditions, procedures and general provisions governing official statistics.
Nothing of this sort has previously existed at EU level – perhaps because up to now it has not really been required."

Statistics mirror European integration

So why didn't the Commission or, indeed, Eurostat, take the initiative before? **Engelage** traces the development to the Treaty on European Union.

Statistics, in his view, not only mirror European integration, they go further: they always have to be one step ahead in providing information as a factual basis for making decisions. And this, of course, is true not just of this specific Treaty, but of each step on the road to European Union.

Demands on official statistics themselves are another important motive. To be trustworthy, statistics have to abide by certain principles: impartiality, relevance, cost-effectiveness, confidentiality and transparency.

Also, the growing contradiction between ever-greater demand for statistical information and dwindling resources calls for making savings and rationalising working methods. "For this reason alone", he says, "it has become increasingly important for EU statistics to define the basic conditions for homogeneous action."

This is exactly what has happened, albeit a trifle late. "The text contains many tried and tested methods of conduct and cooperation developed at a very early stage without any explicit legal basis. These led to an extremely close partnership between national statistical offices and Eurostat, as advanced as any within the EU. These has now been codified and firmly anchored."

A long time

The new Regulation...

- firstly, specifies procedures that must underlie decisions to be taken on various statistical programmes (pluri-annual, annual and specific) and divides responsibilities between national and Community authorities

- secondly, affirms the need for those involved in Community statistical action, at both national and Community level, to adhere to the same fundamental principles to ensure statistics are scientifically independent, transparent, impartial, reliable, pertinent and cost-effective

- finally, contains minimum rules to be complied with to safeguard statistical confidentiality, as defined on a common basis at Community level – and guarantee the transmission of confidential data held by national authorities to Eurostat where necessary to compile statistics for the whole Community.

It has taken a long while to get so far: "There was never a problem", according to **Engelage**, "in getting the message across that we had to set up a common procedure in order to ensure high quality and avoid duplication. We knew this initiative would be welcomed by all.
The problems only really began when, to achieve the aim of creat-

ing a uniform basis, we had to come to terms with 15 different models, none of which would be relinquished easily."

Compromises had to be made. Was it worth it?

Main benefits

Main 'benefits' expected may be summarised as:

- Introduction and definition of the concept of 'Community statistics'

- Reinforcing and safeguarding the role of Eurostat as the 'Community authority in charge of statistics'

- Reinforcing and safeguarding the decision-making procedures (five-year programme, annual work programme etc)

- Safeguarding the role of the Statistical Programme Committee

- Recognition of a legal status for the fundamental principles governing the production of Community statistics
- Recognition of a legal basis for dissemination as a 'part of the production process of Community statistics' - Definition of 'confidential statistical data' at Community level

- Right of access to administrative sources at national and Community level.

Role of Eurostat

As a necessary complement to this Basic Regulation, an internal Commission Decision has been taken on the role of Eurostat in the production of Community statistics. Its main purpose is to clarify the role of the ' Community authority' (Eurostat) defined in the Basic Regulation, and the division of responsibilities between Commission Services participating in production of such information at Community level. This is to ensure the coherence, feasibility and consistency of Community statistics.

The new Regulation only applies to Community statistics. However, given these act as a model for integration and the Regulation reflects the basic principles of official statistics of the United Nations, the international statistical community is not unaffected. "This is particularly true of the central and eastern European countries. Statistics are a natural part of their progress towards the Union," says **Engelage**

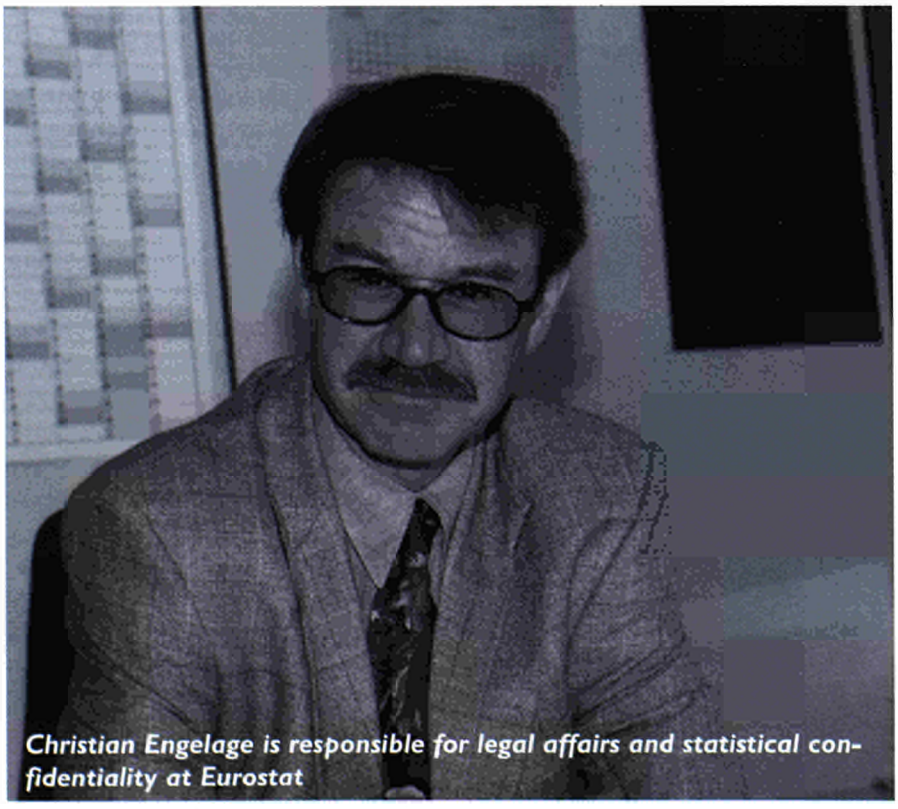
Any interest beyond?

Is this of interest to anyone outside the statistical community? **Engelage** highlights a basic contradiction: "On the one hand, statistics have their own peculiar reputation. Yet there are many areas with a statistical basis that affect every citizen – not least because statistics, despite all their peculiarities, are part of the administrative world.

"So it's an essential precondition of our work that people trust the meaning and purpose of statistics based on clear laws. Only when such trust has been established are people prepared to yield the information required by statisticians, and in reverse are able to deal with the information statisticians can offer. In other words, trust in statistics is indivisible from trust in the state and, by extension, in the European Union.

"This new 'statistical law' creates this transparency. It illustrates and documents how statistics are to be used and codifies the principles of independence, impartiality and confidentiality.

"I am not saying statistics could no longer be produced without this law", was **Christian Engelage**'s opening remark. But clearly he's glad it came to pass.



Christian Engelage is responsible for legal affairs and statistical confidentiality at Eurostat

[1] Council Regulation (EC) N° 322/97 on Community statistics, published in the Official Journal N° L 52 of 22 February, page 1.

Comparing consumer prices...

A story of EU harmony

by Barbara Jakob

In March this year Eurostat published for the first time the Harmonized Indices of Consumer Prices (HICPs). This launch of more comparable indices for the EU and Member States met one condition for the convergence assessment on Monetary Union in the Maastricht Treaty and at the same time marked successful cooperation between Eurostat and national statistical institutes – a study in harmony in more than one way!

Sigma has been following these continuing developments.

In complete harmony – John Astin (middle), head of unit, with his team. From left-to-right: Marie-Noelle Barbier, Per Eckefeldt, Alexander Makaronidis, Daniela Schackis, Angèle Costanzi, Isabelle Bouard and Don Sellwood

The need to harmonize consumer price indices was recognised 20 years ago. But until recently action was restricted to reports reviewing practices in construction of CPIs in Member States.

Eurostat has published Member States' CPIs for several years – simply reproducing national data with no attempt to adjust for methodological differences. The EU CPI published until January 1996 was only a weighted average of unadjusted national indices.

Finally, in 1991, the Maastricht Treaty made the need for harmonization a practical necessity. It laid down various convergence criteria necessary for Economic and Monetary Union. One is a high degree of sustainable convergence in price stability by each country, demonstrated in the inflation rate compared with the three best performing Member States. Inflation has to be measured by comparable consumer price indices. So making Member States' CPIs more comparable was essential.

A common goal

A new delegate working party was established to take the work forward. Its first meeting was in Luxembourg in June 1993 attended by all the then Member States and Austria, Finland, Iceland, Norway and Sweden. There were observers from OECD, ILO and other Directorates-General of the Commission and the Committee of Governors of the Central Banks, all of whom have kept close contact with the project.

Around that time Don Sellwood, a price indices specialist, was seconded to Eurostat from the UK CSO where he had been in charge of the British CPI and related statistics for 20 years. He was therefore well suited to lead the new harmonization project within the unit headed by fellow countryman John Astin. Don Sellwood has now retired but has been persuaded to put his extensive knowledge at Eurostat's disposal as a consultant.

The new working party's task was "to draft guidelines for the construc-

tion of harmonized national consumer price indices suitable for inclusion in a legal instrument". The requirement that "proposals should be practicable for all Member States and should involve minimum cost consistent with the required accuracy of the indices" made it a difficult challenge. Procedure required the working party to agree the texts of Regulations (prepared with the help of task forces) by a qualified majority before putting them to the Statistical Programme Committee (SPC) of EU NSIs' Directors-General for approval.

It was no easy task to reach a consensus so quickly. With different practices in compiling CPIs across Member States, progress of the project depended heavily – and still does – on Member States' willingness to cooperate and compromise. This has been amply demonstrated by the heavy work load it was able to tackle.

In a dynamic economy it is getting more and more difficult to keep the concepts of price indices updated. Problems arise, for example, because of the increasing number of technological innovations and because of the shorter product cycles. The harmonization programme has already confronted many of these critical issues that have recently been raised in the US Senate concerning the accuracy of the US CPI[1].

Step...

In a first step in February 1996, Eurostat launched an interim set of CPIs. This was necessary to provide the Commission and European Monetary Institute (EMI) with data they could use in their first convergence reports to the Council as required by the Maastricht Treaty. These indices were based entirely on existing national CPIs, adjusted only to make product coverage as similar as possible. But they gave a better basis for international comparison than unadjusted national CPIs.

HOW IT'S DONE

The European Index of Consumer Prices (EICP) – or simply the EU average – is calculated as a weighted average of the HICPs of the 15 Member States. The index is computed as an annual chain index allowing for country weights changing each year. The weight of a Member State is its proportion of final consumption expenditure of households in the EU total. The values of final consumption expenditure in national currencies are converted into purchasing power standards (PPS) using the purchasing power parities of final consumption. The country weights used this year are national accounts data for 1995 at 1996 prices. The European Economic Area Index of Consumer Prices (EEAICP) is calculated in the same way including Iceland and Norway.

Certain expenditure categories were excluded where, in the time available, it was impossible to reach agreement on how best to construct comparable measures. In particular, owner occupiers' housing costs, not covered in some countries, measured by equivalent rents in others, and by mortgage interest payments in the rest, were excluded entirely. Spending on health and education was also excluded because of major institutional differences between countries in the ways consumers pay for them, either directly or via taxes. But spending categories not in some national CPIs – in particular, alcoholic drink and tobacco – were included for all Member States.

...by step

HICPs published since March are specifically designed with EMU convergence in mind. Their main purpose is international comparison of consumer price inflation. They are not intended expressly to replace national CPIs which are politically, socially and economically sensitive indicators and cannot readily be changed. HICPs differ from national CPIs in both scope and coverage, as well as in other ways. But, for reasons of economy and practicability, their actual construction depends heavily on that of national CPIs. On the other hand, where improved practices are required for HICPs, their adoption by national CPIs can be expected.

In contrast to the interim indices, HICPs use harmonized methods in many technical areas, not simply coverage. And unlike the interim indices the HICPs also cover insurance for cars and dwellings, package holidays, banking services, evening

Alexander Makaronidis (left) who succeded Don Sellwood (right) as project leader. In the centre is John Astin

language courses and health goods obtainable without prescription.

Efforts were made to ensure that HICPs are up-to-date with market developments. New goods and services should be included in all Member States when they become a significant part of consumption. Furthermore, standards on quality adjustment have been established with the aim of measuring pure price change unaffected by changes in quality or specification.

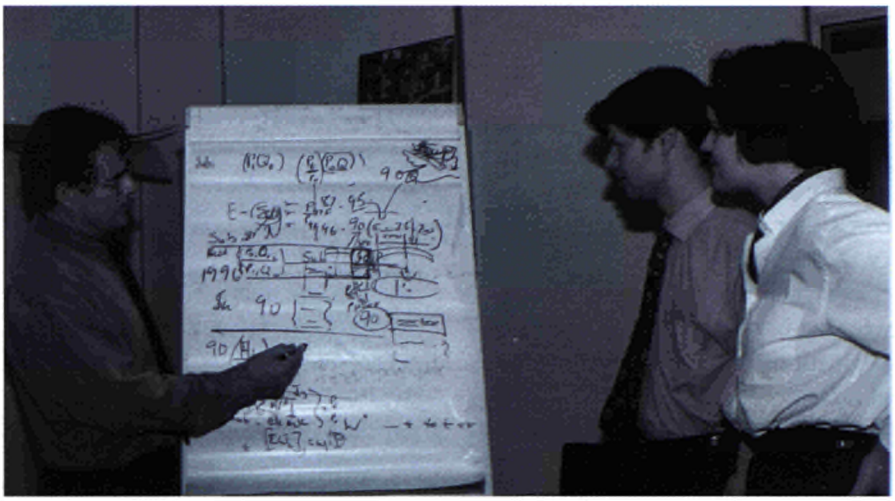
Automatically 'carrying forward' prices, a widespread practice when a particular price is not or cannot be collected, can lead to serious bias. So this is banned in HICPs.

Agreement was also reached on the formula to be used for calculation of 'elementary aggregates', the lowest level of detail for which expenditure weights are known. This was especially difficult as it involved highly technical issues of both concept and practice. A number of Member States will now use a geometric mean formula. The Bureau of Labor Statistics (BLS) is investigating this as well.

The aim of setting minimum standards for sampling was to improve reliability and comparability of HICPs by reducing errors arising from different sample designs and practices.

Proper understanding of inflation requires information on its components. It was agreed Member States should also transmit certain sub-indices for dissemination by Eurostat. For 12 different product groups, which are further divided into several categories, sub-indices are provided by all Member States. These sub-indices – around 100 – are based on a common classification.

HICP working party: Member States proved their willingness to cooperate



Don Sellwood (left) tries to convince his colleagues Per Eckefeldt and Daniela Schackis that measuring inflation is 'really simple'

Tricky points

Some difficult categories such as health and educational services, where there are major institutional differences between Member States, are still not fully covered by the HICPs. A Commission Regulation to tackle such areas is being prepared. There also remains the question of how to measure the impact of inflation on owner-occupiers in respect of housing.

The programme for achieving further harmonization continues. More issues need to be resolved, such as quality of HICP weights, geographical and population coverage, and harmonized treatment of monopolistic prices, such as post, telephones, gas, electricity and water. Once the Euro is in place, the HICP of the participating countries will be used to compile an aggregate for the Euro-zone. This Monetary Union Index of Consumer Prices (MUICP) will be used by, for example, the future European Central Bank.

The outcome of over four years' work is more than a harmonized comparable index of consumer prices. Though there were (and still are) hard discussions on some points and sometimes agreement seems impossible, John Astin and his team also managed to harmonize relations – with and within the NSIs and all parties involved. CPI harmonization is a successful enterprise in more than one respect.

[1] *The so-called* "Boskin Report" *was highly discussed in the media because of its view that the US consumer price index has over-estimated inflation.* Towards a more accurate measure of the cost of living, Final report to the Senate Finance Committee from the Advisory Commission to study the Consumer Price Index, *Michael J Boskin, Chairman, et al, 4 December 1996.*

Legal Acts

Application and implementation of guidelines for constructing the HICP are backed by Community law:

- **Council Regulation (EC) No 2494/95 on harmonized indices of consumer prices, OJ No L 257/1, 27.10.95.**
- **Commission Regulation (EC) No 1749/96 on initial implementing measures for Council Regulation (EC) No 2494/95 on harmonized indices of consumer prices, OJ No L 229/3, 10.9.96.**
- **Commission Regulation (EC) No 2214/96 on the transmission and dissemination of sub-indices of the HICP, OJ No L 296/8, 21.11.96.**

CATHERINE EGINARD *discusses use of administrative sources for statistics and a seminar that focused on this key topic.*

Simple answer to irreversible trend

A Eurostat seminar on use of administrative sources for statistics, held in Luxembourg on 15-16 January, was attended by over 200 experts in the field.

Theme was that use of such sources should be a simple answer to the irreversible trend of statistical development, but there must be minimum conditions for access to administrative information and using it effectively.

The meeting enabled discussion and exchange of experience among many people involved in this area: enterprises and industrial federations, statistical bodies and national and Community authorities.

Member States' national statistical offices are turning more and more to other sources to produce high-quality information at lower cost and ease the response burden on enterprises. So statisticians see use of administrative sources as an alternative or supplement to traditional methods.

Use of data collected for non-statistical purposes leads to certain difficulties of reconciliation. Administrative data collected for purely national purposes do not always satisfy the comparability and quality criteria that are an essential characteristic of the European statistical system.

Seminar discussions revolved around reducing the response burden and cost of surveys; access and confidentiality; quality and comparability of data; and need for cooperation between statistical and administrative authorities.

Reduction of costs

All speakers stressed that use of administrative sources satisfied the current need to reduce the response burden without raising costs to data users. There are reasonable similarities between enterprises and people providing data, on one hand, and statistical and administrative services on the other. Rationalising data collection is one way of reducing costs and easing the statistical burden, and currently constitutes the main and most visible advantage of using administrative sources.

Access to and use of administrative sources is boosted considerably by the existence of a national statistical law. In Germany, for example, there is no general law authorising access to administrative data but specific laws grant rights of access. The new European statistical law (*see article on page 38*) should act as a lever in encouraging statistical use of administrative sources.

Often doubts

It emerged from discussion that there are often doubts about the quality of administrative data: units interviewed are not necessarily the

Yves Franchet, Director-General of Eurostat (centre), with the two chairs of the session, Lidia Barreiros, Eurostat Director (Social and regional statistics and geographical information system) and François de Geuser, Acting Director (Business statistics)

Hans Zeuthen, former Director-General of Denmark Statistics (centre), chairing the discussion panel – from left to right: Pilar Rey (Institute of Fiscal Studies, Spain), Walter Möller (German Federal Ministry of Economic Affairs), Henry Knoop (Tax Manager, Unilever Netherlands), Jean-Pierre Grandjean (Head of Business Statistics System Department, INSEE), Reinhard Schulte-Braucks (Head of DG XXIII-A unit)

right ones for statistical purposes, and there are often differences in regularity and coverage of administrative surveys.

But one speaker pointed that the advantages and disadvantages of using administrative sources followed no set pattern. They depend on each collection. And quality could be improved by combining various sources – for example, administrative data with a sample survey.

Confidentiality stays a thorny issue but legal solutions seek to preserve it. There must be absolute respect for the fundamental principle that data produced by statistical methods aim to be general and aggregated – for statistical use only. Practices in Nordic countries of establishing links between master files are based on trust in statistics.

All presentations showed use of administrative sources is most effective when there is close cooperation between statistical and administrative services. Statistics can then play a role of identification and standardisation that, from the outset, simplifies the collection of administrative data. This helps improve the quality of information and reduce the burden.

Influence of Eurostat

In keeping with the principle of subsidiarity, Eurostat has no direct authority in these matters but does have a fairly strong indirect influence.

For example, it has power to make methodological recommendations in regulations on business registers used for statistical purposes and on structural statistics on enterprises. The seminar felt this should perhaps be developed in the field of social statistics. Eurostat can also act as a forum for Member States to consider this topic and could even study the possibility of technical assistance.

In ensuring proper application of Community legislation, Eurostat could use its executive powers to persuade government authorities to adopt a clearer stance on the merits of using administrative sources for statistical purposes.

Eurostat's role is to ensure that, for a given objective, statistical regulations help minimise the burden; and to be involved in work on promoting simplification under the leadership of DG XXIII (Enterprise policy, distributive trades, tourism and cooperatives). It also has the task of promoting uniform standards within the EU and simplifying its demand for information.

Development of new legal texts could also be considered to achieve 'total quality'. There might be further rationalisation of the European statistical system – introducing an integrated approach and strengthening standards for conducting administrative surveys: on nomenclatures, statistical units, registers and information quality. This approach has already been developed for business statistics.

***Catherine Eginard** is in the Eurostat unit dealing with industry, iron and steel and coordination of surveys on enterprises.*

Venta • Salg • Verkauf • Πωλήσεις • Sales • Vente • Vendita • Verkoop • Venda • Myynti • Försäljning

BELGIQUE/BELGIË

Moniteur belge/Belgisch Staatsblad

Rue de Louvain 40-42/
Leuvenseweg 40-42
B-1000 Bruxelles/Brussel
Tél. (32-2) 552 22 11
Fax (32-2) 511 01 84

Jean De Lannoy

Avenue du Roi 202/
Koningslaan 202
B-1060 Bruxelles/Brussel
Tél. (32-2) 538 51 69
Fax (32-2) 538 08 41
E-mail: jean.de.lannoy@infoboard.be

Librairie européenne/Europese Boekhandel

Rue de la Loi 244/
Wetstraat 244
B-1040 Bruxelles/Brussel
Tél. (32-2) 295 26 39
Fax (32-2) 735 08 60

DANMARK

J. H. Schultz Information A/S

Herstedvang 10-12
DK-2620 Albertslund
Tlf. (45) 43 63 23 00
Fax (45) 43 63 19 69
E-mail: schultz@schultz.dk
URL: www.schultz.dk

DEUTSCHLAND

Bundesanzeiger Verlag

Breite Straße 78-80
Postfach 10 05 34
D-50667 Köln
Tel. (49-221) 20 29-0
Fax (49-221) 20 29 278

GREECE/ΕΛΛΑΔΑ

G.C. Eleftheroudakis SA

International Bookstore
Panepistimiou 17
GR-105 64 Athens
Tel. (30-1) 331 41 80/1/2/3
Fax (30-1) 323 98 21
E-mail: elebooks@netor.gr

ESPAÑA

Mundi Prensa Libros, SA

Castelló, 37
E-28001 Madrid
Tel. (34-1) 431 33 99/431 32 22
Fax (34-1) 575 39 98
E-mail: mundiprensa@tsai.es
URL: www.tsai.es/mprensa

Boletín Oficial del Estado

Trafalgar, 27-29
E-28071 Madrid
Tel. (34-1) 538 22 95 (Libros)/
384 17 15 (Suscripciones)
Fax (34-1) 538 23 49 (Libros)/
384 17 14 (Suscripciones)
URL: www.boe.es

Mundi Prensa Barcelona

Consell de Cent, 391
E-08009 Barcelona
Tel. (34-3) 488 34 92
Fax (34-3) 487 76 59

FRANCE

Journal officiel

Service des publications des CE
26, rue Desaix
F-75727 Paris Cedex 15
Tél. (33-1) 40 58 77 01/31
Fax (33-1) 40 58 77 00

IRELAND

Government Supplies Agency

Publications Section
4-5 Harcourt Road
Dublin 2
Tel. (353-1) 661 31 11
Fax (353-1) 475 27 60

ITALIA

Licosa SpA

Via Duca di Calabria, 1/1
Casella postale 552
I-50125 Firenze
Tel. (39-55) 64 54 15
Fax (39-55) 64 12 57
E-mail: licosa@ftbcc.it
URL: id382.cilea.it/Virtual_Library/bibliot/vetrina/licosa/tt1.htm

GRAND-DUCHÉ DE LUXEMBOURG

Messageries du livre Sarl

5, rue Raiffeisen
L-2411 Luxembourg
Tél. (352) 40 10 20
Fax (352) 490 661
E-mail: mdl@pt.lu

Abonnements:

Messageries Paul Kraus

11, rue Christophe Plantin
L-2339 Luxembourg
Tél. (352) 499 88 88
Fax (352) 499 888 444
E-mail: mpk@pt.lu
URL: www.mpk.lu

NEDERLAND

SDU Servicecentrum Uitgevers

Christoffel Plantijnstraat 2
Postbus 20014
2500 EA 's-Gravenhage
Tel. (31-70) 378 98 80
Fax (31-70) 378 97 83
E-mail: sdu@sdu.nl
URL: www.sdu.nl.

ÖSTERREICH

Manz'sche Verlags- und Universitäts-buchhandlung Gmbh

Siebenbrunnengasse 21
Postfach 1
A-1050 Wien
Tel. (43-1) 53 161 334 / 340
Fax (43-1) 53 161 339
E-mail: auslieferung@manz.co.at
URL: www.austria.EU.net:81/manz

PORTUGAL

Imprensa Nacional-Casa da Moeda, EP

Rua Marquês de Sá da Bandeira, 16 A
P-1050 Lisboa Codex
Tel. (351-1) 353 03 99
Fax (351-1) 353 02 94/384 01 32

Distribuidora de Livros Bertrand Ld.ª

Rua das Terras dos Vales, 4 A
Apartado 60037
P-2701 Amadora Codex
Tel. (351-1) 495 90 50/495 87 87
Fax (351-1) 496 02 55

SUOMI/FINLAND

Akateeminen Kirjakauppa /

Akademiska Bokhandeln
Pohjoisesplanadi 39/
Norra esplanaden 39
PL/PB 128
FIN-00101 Helsinki/Helsingfors
Tel. (358-9) 121 41
Fax (358-9) 121 44 35
E-mail: akatilaus@stockmann.mailnet.fi
URL: booknet.cultnet.fi/aka/index.htm

SVERIGE

BTJ AB

Traktorvägen 11
PO Box 200
S-22100 Lund
Tel. (46-46) 18 00 00
Fax (46-46) 18 01 25
E-mail: btj_tc@mail.btj.se
URL: www.btj.se/media/eu

UNITED KINGDOM

The Stationery Office Ltd (Agency Section)

51, Nine Elms Lane
London SW8 5DR
Tel. (44-171) 873 9090
Fax (44-171) 873 8463
URL: www.the-stationery-office.co.uk

ICELAND

Bokabud Larusar Blöndal

Skólavördustig, 2
IS-101 Reykjavik
Tel. (354) 55 15 650
Fax (354) 55 25 560

NORGE

NIC Info A/S

Østenjoveien 18
Boks 6512 Etterstad
N-0606 Oslo
Tel. (47-22) 97 45 00
Fax (47-22) 97 45 45

SCHWEIZ/SUISSE/SVIZZERA

OSEC

Stampfenbachstraße 85
CH-8035 Zürich
Tel. (41-1) 365 53 15
Fax (41-1) 365 54 11
E-mail: urs.leimbacher@ecs.osec.inet.ch
URL: www.osec.ch

ČESKÁ REPUBLIKA

NIS CR - prodejna

Konviktská 5
CZ-113 57 Praha 1
Tel. (42-2) 24 22 94 33
Fax (42-2) 24 22 94 33
E-mail: nkposp@dec.nis.cz
URL: www.nis.cz

CYPRUS

Cyprus Chamber Of Commerce & Industry

38, Grivas Digenis Ave
Mail orders:
PO Box 1455
CY-1509 Nicosia
Tel. (357-2) 44 95 00/46 23 12
Fax (357-2) 361 044
E-mail: cy1691_eic_cyprus@vans.infonet.com

MAGYARORSZÁG

Euro Info Service

Európa Ház
Margitsziget
PO Box 475
H-1396 Budapest 62
Tel. (36-1) 11 16 061/11 16 216
Fax (36-1) 302 50 35
E-mail: euroinfo@mail.matav.hu
URL: www.euroinfo.hu/index.htm

MALTA

Miller Distributors Ltd

Malta International Airport
PO Box 25
LQA 05 Malta
Tel. (356) 66 44 88
Fax (356) 67 67 99

POLSKA

Ars Polona

Krakowskie Przedmiescie 7
Skr. pocztowa 1001
PL-00-950 Warszawa
Tel. (48-2) 26 12 01
Fax (48-2) 26 62 40

TÜRKIYE

Dünya Infotel A.S.

Istiklál Cad. No: 469
TR-80050 Tünel-Istanbul
Tel. (90-212) 251 91 96
(90-312) 427 02 10
Fax (90-212) 251 91 97

BÅLGARIJA

Europress-Euromedia Ltd

59, Bld Vitosha
BG-1000 Sofia
Tel. (359-2) 80 46 41
Fax (359-2) 80 45 41

HRVATSKA

Mediatrade Ltd

Pavla Hatza 1
HR-10000 Zagreb
Tel. (385-1) 43 03 92
Fax (385-1) 44 40 59

ROMÂNIA

Euromedia

Str. G-ral Berthelot Nr 41
RO-70749 Bucuresti
Tel. (40-1) 210 44 01/614 06 64
Fax (40-1) 210 44 01/312 96 46

SLOVAKIA

Slovenska Technicka Kniznica

Námestie slobody 19
SLO-81223 Bratislava 1
Tel. (42-7) 53 18 364
Fax (42-7) 53 18 364
E-mail: europ@tbb1.sltk.stuba.sk

SLOVENIA

Gospodarski Vestnik

Zalozniska skupina d.d.
Dunajska cesta 5
SI-1000 Ljubljana
Tel. (386) 61 133 03 54
Fax (386) 61 133 91 28
E-mail: belicd@gvestnik.si
URL: www.gvestnik.si

ISRAEL

R.O.Y. International

17, Shimon Hatarssi Street
PO Box 13056
61130 Tel Aviv
Tel. (972-3) 546 14 23
Fax (972-3) 546 14 42
E-mail: royil@netvision.net.il

Sub-agent for the Palestinian Authority:

Index Information Services

PO Box 19502
Jerusalem
Tel. (972-2) 27 16 34
Fax (972-2) 27 12 19

RUSSIA

CCEC

60-letiya Oktyabrya Av. 9
117312 Moscow
Tel. (095) 135 52 27
Fax (095) 135 52 27

AUSTRALIA

Hunter Publications

PO Box 404
3167 Abbotsford, Victoria
Tel. (61-3) 9417 53 61
Fax (61-3) 9419 71 54

CANADA

Uniquement abonnements/
Subscriptions only:

Renouf Publishing Co. Ltd

1294 Algoma Road
K1B 3W8 Ottawa, Ontario
Tel. (1-613) 741 73 33
Fax (1-613) 741 54 39
E-mail: renouf@fox.nstn.ca
URL: fox.NSTN.Ca/~renouf

EGYPT

The Middle East Observer

41, Sherif Street
Cairo
Tel. (20-2) 39 39 732
Fax (20-2) 39 39 732

JAPAN

PSI-Japan

Asahi Sanbancho Plaza #206
7-1 Sanbancho, Chiyoda-ku
Tokyo 102
Tel. (81-3) 3234 69 21
Fax (81-3) 3234 69 15
E-mail: psijapan@gol.com
URL: www.psi-japan.com

SOUTH AFRICA

Safto

5th Floor Export House,
CNR Maude & West Streets
PO Box 782 706
2146 Sandton
Tel. (27-11) 883 37 37
Fax (27-11) 883 65 69

UNITED STATES OF AMERICA

Bernan Associates

4611-F Assembly Drive
MD20706 Lanham
Tel. (301) 459 2255 (toll free telephone)
Fax (800) 865 3450 (toll free fax)
E-mail: query@beman.com
URL: www.beman.com

MÉXICO

Mundi-Prensa Mexico, SA de CV

Río Pánuco, 141
Delegación Cuauhtémoc
ME-06500 México DF
Tel. (52-5) 553 56 58/60
Fax (52-5) 514 67 99
E-mail: 104164.23compuserve.com

RÉPUBLIQUE DE CORÉE

Kyowa Book Company

1 F1. Phyung Hwa Bldg
411-2 Hap Jeong Dong, Mapo Ku
121-220 Seoul
Tél. (82-2) 322 6780/1
Fax (82-2) 322 6782
E-mail: kyowa2@ktnet.co.kr.

ANDERE LÄNDER/OTHER COUNTRIES/ AUTRES PAYS

Bitte wenden Sie sich an ein Büro Ihrer Wahl / Please contact the sales office of your choice / Veuillez vous adresser au bureau de vente de votre choix

E U R 1 5 Prodcom/Combined Nomenclature

E U R O P R O M S

European production and market statistics

1st edition, 1997

UNIQUE TOOL FOR ANALYSING EU MARKETS

For the first time Eurostat is offering a **CD-ROM** with detailed and comparable data on production and external trade. It enables analysis of the domestic markets of almost **4,400 industrial products in the EU.**

Issued in the context of the European statistical system, the **CD-ROM – EUROPROMS –** includes sectors as diverse as mining, agri-foodstuffs, intermediate products and capital and consumer goods.

Data cover **1993, 1994, 1995** and first two quarters of **1996.**

EUROPROMS is a unique tool for manufacturers, multinational firms, small and medium enterprises, consultants, academics – all who need to compare sectors of EU industry and understand their current and future position.

SYSTEM REQUIREMENTS:

- PC with 486-33 Mhz processor (minimal) or Pentium 100 Mhz (recommended)
- 8 MB RAM (minimal), 16 MB (recommended)
- Double speed CD-drive (minimal), 4X (recommended)
- 15 MB free harddisk space
- 256-colour Super VGA display (640 x 480 resolution)
- Windows 3.1 or Windows 95 operating system

Catalogue n°: CA-03-97-822-4H-Z
Price in Luxembourg, excluding VAT: ECU 2 000; Library price, excluding VAT: ECU 1 500
4 language version (ES/EN/DE/FR)

Data Shop Luxembourg • 2, Rue Jean Engling • L-1466 LUXEMBOURG • Tel.: (352) 43 35 22 51 •Fax: (352) 43 35 22 221
Data Shop Bruxelles • Rue de la Loi, 130 • B-1049 BRUSSELS • Tel.: (32-2) 299 66 66 • Fax: (32-2) 295 01 25

OFFICE FOR OFFICIAL PUBLICATIONS
OF THE EUROPEAN COMMUNITIES

L-2985 Luxembourg

STATISTICS
EUROPE

CA-AB-97-002-EN-C

ISSN 1018-5739