

# Research and Development in Statistics in the European R&D Framework Programmes

Jean-Louis Mercy, Photis Nanopoulos, Harald Sonnberger

*European Commission*

*Eurostat*

*L – 2920 Luxembourg*

[jean-louis.mercy@cec.eu.int](mailto:jean-louis.mercy@cec.eu.int), [photis.nanopoulos@cec.eu.int](mailto:photis.nanopoulos@cec.eu.int), [harald.sonnberger@cec.eu.int](mailto:harald.sonnberger@cec.eu.int)

<http://europa.eu.int/comm/eurostat/research/>

## **Résumé**

*Afin de faire face aux défis toujours plus importants que présente la production de statistiques harmonisées au niveau européen, Eurostat - l'office statistique de l'Union Européenne – promeut le développement et l'utilisation de nouvelles méthodes et techniques. Pour ce faire, différentes initiatives ont été lancées, dont la plus importante est le financement d'activités de recherche et développement en statistique, au travers des programmes cadres de recherche de l'Union Européenne. Quantitativement, au cours des 6 dernières années, environ 80M€ ont été consacrés au financement de 150 projets, regroupant 500 participations de centres de recherche, du monde industriel et des instituts nationaux de statistiques. L'objectif de cette contribution est de décrire les activités de R&D en statistique financées par Eurostat, de tirer quelques conclusions sur l'impact de cette action et d'esquisser son avenir dans le cadre du nouveau programme de recherche de l'Union Européenne.*

## **1. Introduction**

The challenges for the European Statistical System (ESS – network of national and international public bodies in charge of providing the statistical information needed by the E.U. policies) have never been greater than they are at present. New phenomena, such as globalisation, the appearance of the knowledge society and economy, the effects of the “new economy” on employment and competitiveness, the creation of and the need for short-term indicators for the euro-zone, the on-going and up-coming negotiations for EU membership with the candidate countries, the new political need for statistical indicators used in the context of the benchmarking method, etc. have all increased the pressure on the ESS and in particular on Eurostat (the statistical office of the European Union), to produce relevant, timely, and accurate statistical information.

To face these challenges, since 1989, Eurostat has promoted the development and the use of new methods and technologies, as well as the reinforcement of co-operation between the ESS, the scientific communities and the international area.. The main mechanisms to implement this activity have been the European Union R&D Framework Programmes, in which R&D in statistics has steadily grown in importance. Eurostat has also a strong tradition in the organisation of scientific conferences (e.g. KESDA 98, NTTS 98, ETK 99, ETK-NTTS' 01) and Workshops (e.g. Statistical Metadata, Confidentiality), in which areas of particular relevance for Official Statistics are discussed together with academics and representatives of the private sector. Furthermore, Eurostat publishes the ROS journal (Research in official statistics), which presents papers of high scientific content resulting from research carried out in the field of official statistics, and relevant to the work of the statistical offices.

## **2. The past research activities**

During the second Framework Programme, a first action (DOSES – Development of statistical expert

systems - 1989-1992) involved the funding of 7 projects, for a total budget of 4 M€ This relatively small programme was designed to support co-operation between academic and official statisticians, between universities and the private sector in the development of marketable software, and to stimulate the international dimension of research. DOSES highlighted the need to introduce modern information technology for the processing of public and official statistics, and the need for harmonisation and creation of interfaces between national standards.

DOSIS (Development of statistical information systems – 1995-1998) developed during the fourth Framework Programme, strongly emphasized on the application of results for improving the competitive position of European industry and the efficiency of National Statistical Institutes. The Framework Programme was also designed to encourage co-operation through the creation of transnational consortia drawn from academia, government and the private sector. DOSIS co-financed 18 projects for a total budget of 17.6 M € More than half of the DOSIS projects have finished. Overall, they were more concentrated on the technological aspects of the statistical process: automated data collection, integration of databases, metadata processing, disclosure control. Other important topics include data analysis and knowledge extraction.

At the same time, in another budget context of the fourth Framework Programme, an initiative was undertaken to fulfil the research needs of the Commission services, and was used by Eurostat to provide and disseminate existing information and to harmonise the system of data collection and analysis which was in place in the European Union. This programme, called Sup.Com, funded 93 projects of direct interest to Eurostat for a total of 13.7M€ Half of the Sup.Com projects concentrated on methodological issues, dealing with harmonisation and comparability (often applied to social statistics), environment, transport and business. The adaptation of the statistical system to the new IT realities (e.g. use of the Internet, new graphical tools, computer assisted training, etc.) was also covered. Emerging phenomena and techniques were explored (e-com, neural networks etc). The first results of an on-going external evaluation show that 90% of the projects evaluated contribute to the work of Eurostat.

### 3. The current programme

The fifth Framework Programme, which was initiated in 1998, is the first to recognise explicitly the importance of statistics for the information society. Research in statistics was included as a specific action line, after consultation with policy makers, National Statistical Institutes, research communities, and the private sector.

EPROS (European Plan of Research in Official Statistics) was drafted for the implementation of the fifth Framework Programme and in response to the challenges of the ESS. The guiding principles of EPROS are that projects should have clear and appreciable R&D content, reflect the needs of the ESS and other producers/users of statistics, make use of new developments in technology, have a clear exploitation plan, and rest on cross-national and multidisciplinary partnerships.

The implementation of EPROS involves many groups of actors: official and academic statisticians with theoretical and domain expertise; computer scientists and information managers; other professionals, e.g., economists, in multidisciplinary teams; users, producers and providers of data and information technologies. During its first two years of implementation, 35 projects have been financed, with approximately 50M€

#### *Five key components of EPROS*

- **Research & Development into generic statistical tools, techniques, methodologies and technologies:** this covers the entire statistical production process, with particular emphasis on achieving the highest standards of data quality, and on automating data collection, capture, interchange, editing, imputation and dissemination.

- **Statistical Applications:** this component addresses the use of statistical tools, techniques, methods and technologies in support of domain research activities in other parts of the research programmes.
- **Statistical Indicators for the New Economy (SINE):** recent trends in Information and Communications Technology (ICT) together with the evolution of globalisation have led to the emergence of the New Economy. The long-term structural shift from the industrial economy towards the services and the information economy must be monitored by statistical indicators. It is likely that conventional concepts and classifications will be inappropriate. Hence, there should be considerable emphasis on innovative, forward-looking indicators relating to emerging activities.
- **Transfer of technology and know-how initiative (TTK):** a key objective of the R&D activities is to achieve transfer of technology and know-how in the ESS and other relevant organisations and institutions that could benefit from the knowledge and tools developed through EU research, either steered by the Commission, by national R&D programmes or by NSIs themselves.
- **Supporting activities:** other complementary measures are foreseen to prepare, extend or disseminate the R&D and TTK projects.

The first projects funded under the fifth Framework Programme started in 2000. A large range of activities is covered, such as: improvement of classifications, automated questionnaires, use of new data sources, ex-post harmonisation techniques, validation, editing and imputation; estimation, modelling and analysis (business cycle, small area, etc.), metadata (with a specific interest on standardisation), disclosure control. The statistical indicators of the New Economy are addressed via the description of the content of the Internet, the tracking of Ecom, new indicators for the knowledge economy, innovation through patent databases, improvement of statistical information systems, etc.

#### 4. Conclusion

From a quantitative viewpoint, the R&D activity undertaken or supported by Eurostat is considerable, and fits with the objectives. In the last 6 years, 80 M€ have been spent on approximately 150 projects, involving close to 500 research teams, covering the 15 Member States. 130 have had NSI participation from 12 Member States. The aims of promoting research in the ESS, of reinforcing the research potential, of strengthening relations between official and academic statisticians and of reinforcing international co-operation in R&D in statistics are gradually being reached.

From a qualitative viewpoint, due to the diversity of the research domains and projects, it is difficult to draw general conclusions. The exploitation of project results includes pre-commercialisation of software, inclusion of new tools in the production and dissemination environment in the ESS, dissemination of products through the Internet, scientific publications, and recommendations of new methods to use. Descriptions and results of individual projects are progressively made available via the Eurostat web site<sup>1</sup>.

The research programme has now reached a stage where results and experiences from earlier projects lead to an accumulation of know-how. Simultaneously, research on specific themes is no longer dependent on a single project. Rather the integration of work from several projects is feasible. Two examples, metadata and data quality, are given below.

**Metadata.** In the fourth Framework Programme, several projects had a strong statistical metadata content. This has continued in the fifth Framework Programme. Projects with a metadata component such as IMIM, IDARESA, ADDSIA, FASTER, IPIS, IQML, MISSION, METAWARE (see the web site<sup>1</sup> for details) belong to this group. At the same time, several international bodies are dealing with the harmonisation of metadata; a need for convergence in the metadata area has been identified.

With the critical mass coming from these projects, a new type of project, a thematic network called

---

<sup>1</sup> <http://europa.eu.int/comm/eurostat/research/index.htm>

METANET, is now developing integrating proposals for views and common standards and disseminating the results. Another project, COSMOS, was recently initialised. Its purpose is to identify metadata objects common to several projects and to implement them.

**Data quality and best practices.** Several projects address the issues of data quality, specifically the areas of imputation, editing and (variance) estimation. The projects AUTIMP (fourth Framework Programme), EUREEDIT, EURAREA and DACSEIS address new and innovative aspects of these classical practices in official statistics. A common approach is the evaluation of methods, where the performance of existing and new methods are compared empirically, on real datasets or in simulated but realistic populations. Best practices and comparisons of methods are also addressed in the area of statistical confidentiality (CASC), in harmonisation (CHINTEX) and various other fields. It is foreseen that the projects will both cluster among themselves and contribute to general 'best practices' guidelines in the data quality area.

Several other themes such as data capture, data integration, time-series techniques, automated coding, and EDI have emerged from the present programme. It is expected that in the forthcoming years many of these themes will reach a critical mass and, through consolidation, will lead to a considerable contribution to the development of the European Statistical System.

Since the exploitation of results and the use of newly developed products must be improved, Eurostat has engaged in a more proactive dissemination, exploitation and transfer of technology initiative, in order to reinforce the capacities of the ESS.

International co-operation in R&D has been strengthened inside the E.U. Due to the globalisation and the need for more comparability world-wide, a co-operation agreement has been signed between Eurostat and the United States Office of Management and Budget, for the development of co-operative activities in the field of R&D in statistics. The agreement covers several domains of application such as the statistics on Science, Technology and Innovation, the New Economy, the information society. The need for co-operation in quality measurement, confidentiality, time series training, IT tools, knowledge warehousing and management is also identified. Such bilateral agreements could be repeated with other countries.

## **5. Future prospects**

The present EPROS programme will continue in the annual workprogrammes of the IST programme until 2002.

First steps to plan future research beyond the present fifth Framework Programme have been taken. The European Commission has made a proposal for the sixth Framework Programme in February 2001. In the proposal, statistics and statistical indicators are mentioned in several places. The questions of management and implementation of the programme are not yet decided.

Meanwhile, Eurostat has launched the planning of the contents of the Next Framework Programme, by inviting a group of international experts to produce a proposal for a research programme in official statistics. A first vision paper was introduced in December 2000 to the representatives of the Member States. Eurostat has sent out a second, more detailed version for consultation in the spring of 2001. The second version contains a questionnaire, and comments from the National Statistical Institutes, the scientific community and the industry are expected in the first half of the year 2001. A final research plan will be drafted based on this consultation.