# Developments in EU Statistics on Science, Technology and Innovation: taking stock and moving towards evidence based policy analysis

**Prof. Dr. Reinhilde Veugelers**
**EC-BEPA & Katholieke Universiteit Leuven & CEPR, London**

*Paper prepared for the Blue Sky Indicator Conference, Ottawa, September 2006, Statistics Canada*[1]

## 1. Introduction

The Lisbon European Council in 2000 rightly recognised that Europe's future economic development would depend crucially on its ability to create and grow high value, innovative and research-based sectors. EU and national action plans have accordingly been developed. In view of the scope of the problem, it was clear that tackling the deficient EU innovative capacity required a longer-term, broad, systemic policy framework. No single action would deliver innovation based higher growth. Rather, a series of interconnected initiatives and structural changes were needed. Going beyond stimulating the research inputs from the public and the private sector, other structural reforms in product, labour and capital markets became part of the policy agenda as well. Nevertheless, the EU's R&D deficit became a central focus of research policy with the articulation of the 3% target for R&D expenditures at the Barcelona European Council in 2003.

Years later, Europe's performance in R&D, innovation and growth continues to be a disappointing story. While there are examples of good performance in particular sectors and particular Member States, overall the EU innovation environment remains weak. If recent trends continue, nor the 2% private nor the 1% public target will be reached by 2010.

With lacking results comes fatigue, lack of interests and mounting criticism on policy: not enough public funding dedicated to innovation; lack of governance; no real commitment beyond rhetoric; wrong and missing instruments. Evaluations of EU's innovation policy (such as the Kok Report 2004, Aho Report (2006)) have indicated as major problem the governance of innovation policy, including the need for better informed policy makers.

The way forward for improving innovation policy in Europe is in better analysis and diagnosis to guide ***evidence based policy design***. Policies need to be supported by analysis, monitoring and evaluation practices, which then feed back into the policy process. This note examines the challenges for the statistical system delivering in an

[1] The paper draws heavily on a note produced for EC-Commissioner's Potocnik's Knowledge for Growth Working Party on S&T Indicators ("Current situation and recent developments of the EU statistics on science, technology and innovation" by V. Duchene, A. Gotzfried & R.Veugelers, 2006)

evidence based policy process. To some extent, the current official statistics have progressed, but however show deficiencies with regard to the measuring of the knowledge economy and feeding into policy analysis.

The adoption of the EU legal framework for Science, Technology and Innovation statistics in 2003/2004 constituted a major step forward because the mandatory provision of basic statistics following a common methodology is the basis for EU policy learning and policy development. However, the statistics defined in the legislation and the underlying methodology is limited: for R&D statistics it is based on the Frascati approach to R&D, which is essentially an input approach, so lacking basic features of a systemic framework for policy learning and development.

Over the past ten years, some experiments with ‘process indicators’ (describing ‘knowledge flows’ or ‘research networks’) and ‘micro-level indicators’ (monitoring the behaviour and performance of key actors such as firms, universities), have tried to remediate this problem. These pilots, however, were conducted on ‘ad-hoc basis and did not yet result in internationally comparable data and methodologies. Therefore, at this moment, the mandatory provision of statistical data according to the current EU legal regulation is still the most reliable source for policy monitoring.

The aim of this note is to indicate where the main gaps are in the current Statistics on Science, Technology and Innovation, and to propose ways of improving the current situation.

The note is structured as follows.

- What data/indicators do we need (section 2)
- Taking stock of indicators used in the EU policy process (section 3)
- Evaluating the choice of indicators (section 4)
- The way forward (section 5)
- Conclusions

## 2. What data/indicators do we need?

Starting from a policy perspective on how to improve EU's growth, it is important to broaden the scope of the exercise beyond data and indicators in the area of R&D only. Taking the perspective of assessing "***innovation capacity***" defined as the ability of a system to not only produce new ideas but also to commercialize a flow of innovative technologies over the longer term, a range of factors are deemed to be important for effective innovation effort. A sufficiently developed 'supply' side of R&D (as reflected in the amount of R&D carried out or the number of skilled researchers) is a necessary but insufficient condition for successful innovation. Broader framework conditions are important as well, including a sufficient 'demand' for innovation to reward successful innovators. A critical element in the framework is the interconnectedness of the agents in the system, linking the common innovation infrastructure to specific technology clusters. This requires a.o. good industry-science links and well functioning capital and labour markets.

In the Innovation Capacity perspective, country differences with respect to innovation and growth might reflect not just different endowments in terms of labour, capital and the stock of knowledge, but also the varying degrees of the "knowledge distribution power" or the efficiency of the innovation system. Overall, this perspective warns against (1) looking at knowledge creation indicators only and (2) looking at statistical indicators individually to assess performance. Rather, a **systemic approach** should be taken to understand the relationships between STI and socio-economic development.

The problem with this approach, however, is to approximate empirically the institutional framework and the "knowledge distribution power" of nations. What is available at present are only pieces of statistical evidence showing the importance of interactions, such as the availability of venture-backed financing, cooperation in R&D among firms and between science and industry, (international) co-patenting, the number of researchers employed by business,…

# 3. Taking stock of indicators currently used in the EU policy process

The wide scope of the Lisbon strategy made it necessary to focus and identify a restricted well defined set of targets and policy measures necessary to achieve the objectives, and at the same time, a corresponding restricted set of indicators to monitor progress on the targets.

The selection of the indicators is clearly constrained by the availability of data: i.e. series should be available/ comparable across Member States and with major benchmarks (like the US, Japan, China…) and over time and sufficiently recent.

## *(i) Structural Indicators*

To monitor the progress on the main targets of the Lisbon strategy, the Commission and the Council agreed on a list of 14 main structural indicators. Member States' performance on these indicators is continuously being assessed.

For the European Knowledge Area, R&D expenditures as a % of GDP, with a target of 3% is the main indicator. But beyond this main indicator there are other (secondary) structural indicators of the EKA, which are continuously being monitored:

- Spending on human resources (public expenditure on education)
- GERD (Gross domestic expenditure on R&D) by source of funds (private-public)
- Level of Internet access – households /enterprises
- Science and technology graduates – total/females/males
- Patents – EPO/USPTO
- Venture capital investments – early stage/expansion & replacement
- ICT expenditure – IT/Telecommunications expenditure

## *(ii) Innovation Indicators*

Beyond the structural indicators which cover all Lisbon areas, the Lisbon European Council also requested for the area of innovation and R&D, the development of the European Innovation Scoreboard (EIS) by DG Enterprise. The EIS contains indicators, selected to summarize the main drivers and outputs of innovation. These indicators are divided into four groups: *Human resources for innovation* (S&E graduates, Population with tertiary education, Employment in high-tech…); the *Creation of new knowledge* (Public and Private R&D, EPO & USPTO Patents); the *Transmission and application of knowledge* (Innovation Expenditures, SMEs innovating & cooperating); and *Innovation finance, output and markets* (high-tech & early stage venture capital, sales of innovative products, internet access, ICT expenditures). The EIS mainly uses Eurostat data, covering 32 countries. A number of indicators are drawn from the EU Structural Indicators.

### *(iii) Research Indicators*

In support of the European Research Area initiative (ERA), DG Research was entrusted with a mission to produce a set of indicators and a methodology for benchmarking research policies in the Member States. A set of indicators was proposed to help monitor and report on progress towards the ERA. Most of the indicators are already used in other Commission publications. (eg Key Figures).

## 4. Evaluating the choice of indicators

### *(i) Are we measuring the right indicators for informing improvements in innovative capacity?*

The set of indicators -both the structural indicators, and those specific for research and innovation- although being restricted by data availability, clearly look like being inspired by the specific weaknesses of the EU innovative capacity and the 'systems' approach towards improving this capacity. Although R&D spending is a central **structural indicator**, it fits into a set of other indicators allowing integration with labour, capital, and product market reforms.

A key message from a systemic approach is that the effectiveness of innovation systems depends on the balanced combination of creative capacity, diffusion capacity and absorption capacity. These dimensions are all somehow reflected in the selected indicators. The targets selected for the European Knowledge Area, beyond R&D expenditures, reflect the importance of a highly educated labour force as central in EU's creative but also distributive capacity. It also reflects the specific importance of ICT in the EU's growth agenda as a General Purpose Technology and recognizes the importance of financing for innovation.

Nevertheless, there are some important areas not covered:

### A. Linking Indicators

The area of indicators that is least represented is Scientific Performance and Industry Science Links. Especially the lack of Industry Science Link Indicators is disturbing since this is one of the particular deficiencies of the EU innovative capacity (cf European Paradox). What we are missing in the set of main indicators are those on Industry-Science Links, such as for instance, cooperation between firms and research institutes, co-patenting & co-publishing, researcher mobility between industry and science, private research funding of basic research, patenting by universities and public research institutes, spin-offs…This is partly due to a lack of systematic data on this, but clearly more could be done here (cf infra).

### B. International dimension: Connecting the National/EU innovation capacity to the global system.

The dominant perspective of innovation capacity indicator collection and monitoring exercises is national: trying to assess the activities going on within the national territory. But clearly, economic agents, and particularly large Multinational Firms, who drive the innovation potential of nations, are increasingly operating cross-nationally, with important implications for the performance of national innovation systems. The implication for a national innovation system is that (i) relevant actors include foreign actors operating within or linked to their national innovation system and (ii) national actors are operating in or linked to other innovation systems; Both of these are likely to have an impact on the performance of the national system. We are currently lacking indicators tracing the international activities of actors and their links to national innovation systems.

## *(ii) Are we measuring the indicators at the right level of aggregation?*

### A. Sectoral/technology level

Underlying any aggregate innovation indicator, is the structural make up of the economy, which differs greatly between EU countries. Such structural differences can have an important role in explaining some of the differences in innovation performance. The main reason is that there is a great deal of diversity amongst industrial sectors and technologies in terms of innovation process, innovation inputs and outputs. Technological opportunities differ across sectors with ICT as a prime example of the high growth sector with huge opportunities for technological advance. Other major differences across sectors are the size of the innovating unit, the ease of and methods of appropriation, the motives for innovation, with product versus process innovations. There is also a great deal of diversity amongst the sources of innovation ranging from in-house R&D laboratories over supplier, users, to public research institutes for the industries using science-based technologies. This implies that there will be major differences across sectors in many of the indicators used. The single most important constraint is the lack of data at the sector/technology level for some key variables.

### B. Regional dimension

Another area of over-aggregation is the geographical dimension. Innovation policies are often developed and implemented at the regional level, in addition to national and EU level policies. Regional indicators can help inform these policies. For instance, many innovative activities are strongly localized into clusters of innovative firms, sometimes in close co-operation with public institutions such as research institutes and universities. The effective design and implementation of cluster policies therefore depend on identifying both highly innovative regions and less innovative regions that might have future potential

### C. Micro-level: assessing the behaviour of individual actors: individual researchers, firms, research institutes…

S&T indicators need to measure a wider range of actors (universities, research councils, transfer platforms, start ups, multinationals, venture capital firms, etc). Understanding what drives their behaviour is an important component towards understanding the performance of the aggregate system. Hence, more information is needed at the level of individual actors on relevant parameters (competencies, motives, performance etc) and relationships (financial but also in terms of flows of personnel, strategic coordination etc).

# 5. The way forward

We are still far away for a smooth process of evidence based policy analysis. Work to be done in future includes (i) improving the basic indicators for innovation input and output; (ii) develop new indicators; (iii) disaggregate the data reporting at sectoral/technology; geography and institutional level).

The following section details the current work going on in these areas.

## *(I) Improving Basic Indicators*

### A. R&D expenditure data: Monitoring the contribution of the business enterprise sector to the financing of R&D (Barcelona '2/3 objective')

According to the Frascati Manual and the Commission Regulation 753/2004, expenditure on R&D can be broken down and analysed by sector of performance (sector where R&D activities are executed), and for each sector of performance by source of funding (sector financing the execution of R&D). The 'Barcelona 3% Objective' (3% of GDP should be devoted to R&D) refers to the execution of R&D, while the 'Barcelona 2/3 Objective' (2/3 of total expenditure on R&D should be funded by business enterprise sector) refers to the source of funding[2].

Up to now the monitoring of the so-called 'Barcelona 2/3 objective' has not been complete since part of the funding from the private sector is contained in 'R&D funded from abroad'. There is need to push national statistical offices to get a more complete matrix, and to try to make these data mandatory under the next revision of the regulation (from 2007 onwards).

2 Total R&D expenditure (GERD) can be broken down into four performing sectors: 1) business enterprise ('Business Enterprise Expenditure on R&D' -BERD), 2) government (Government intramural expenditure on R&D –GOVERD), 3) higher education (Higher education expenditure on R&D –HERD) and 4) private non-profit (PNPRD). GERD can also be broken down by source of fund into four funding sources: 1) business enterprise, 2) government, 3) Other national sources, 4) Abroad (Frascati Manual 2002).

## B. Public expenditure on R&D and the 'Government Budget Appropriations or Outlays on R&D' (GBAORD)

When policy analysts monitor the 'public funding of R&D' (the so-called '1/3 Objective'), the indicator that matters is the 'GERD (=total domestic R&D expenditure) financed by government'. This indicator is computed 'ex post' based on statistics delivered under the Commission Regulation 753/2004 or OECD data for non EU countries. Next to this information source that typically comes available for analysis with a considerable delay, GBOARD data provide information on public *budgets* on R&D that is available on shorter notice. GBOARD data are therefore an interesting source to monitor public efforts on R&D. These statistics (GBOARD versus GERD funded by Govt) are commonly used in a complementary manner, although they differ in two fundamental ways:

- GBOARD provides public budget figures, and no expenditure data;
- while 'GERD funded by Govt' is purely domestic, GBOARD is not (part of GBOARD may go to international co-operation such as ESA or Airbus programmes).

The correlation between the two series (per country, across time) is very weak. In order to do improve their correspondence, better and internationally better harmonised procedures for labelling R&D expenditures need to be agreed. More work can be done to harmonise the extensive policy information contained in distributions of the R&D budget (GBAORD) according institutional or thematic categories (starting with a revision and better use of the NABS classification breaking down the GBOARD figures into socio-economic objectives ). This would allow inter alia to analyse the allocation of public funding to more basic versus applied research;

## C. Scientific output: scientific publications, citations, impacts

Bibliometric statistics on publications and citations are analysed as a proxy for scientific productivity. For the time being, the most highly-recognised databases, relating to scientific publication and citation data are owned by Thomson ISI. There are some drawbacks and limitations with regard to the sampling frames of the journals and papers but there is no other robust source of information at world level, particularly wrt citations. The nature of the private ownership of these data has been a hindrance to its being mainstreamed.

## D. Technological Output: patents

In 2004/2005 the international Task Force on patent statistics (where various international organisations as well as DG RTD and ESTAT are represented) were entrusted with the objective to establish an internationally harmonised raw data source on patents: PATSTAT. The EPO committed itself to deliver twice per year a harmonised set of raw data which is then to be used for data production by the OECD, Eurostat and others. This effort led already to an increased data production at DG

ESTAT which was put in place and disseminated in late 2005. In addition, DG ESTAT put considerable efforts into the definition of a method of the cleaning of names of patent applicants. The PATSTAT raw data file can be further exploited, by building in linkages to other datasets.

### E. Researchers, Human Resources

Broadly, information on researchers encompasses three layers: the number of researchers and their variations (stocks), their mobility pattern (flows) and why they are moving (motivational aspects and career paths). At the European level, the information on **researcher stocks** is collected through various channels: a) the *R&D statistics* by Eurostat/ OECD; b) the *Community Labour Force survey*; c) surveys on *doctorate holders***,** widely spread among European countries.With regard to the **mobility of researchers**, there is no reliable data source available at this stage. The most often used proxy is the doctorate mobility, based on education data. Fundamental information such as on net inflows or outflows of researchers is therefore not available at international level. There is even less data on the **career paths** of researchers. **The Career of Doctorate Holders (CDH)** statistics are internationally coordinated statistics (involving Eurostat, OECD and UNESCO) that builds on existing work in countries. Surveys on doctorate holders currently exist in several countries but with various objectives, populations (doctorates, graduates etc) and frequency. As a consequence, information is not currently useable for international comparisons. Following the needs expressed by users, a process was set in motion to develop an internationally standardized CDH survey. In addition to financial support to countries, Eurostat will propose a **legal basis** for this survey (as for the R&D survey) which would mean that it would be mandatory for all European countries to provide the data.

## *(ii) Developing New Indicators*

### A. 'Throughput' / 'process' indicators: Knowledge flows, S&T Linkages, University – Industry linkages, Technology Transfer etc

Beyond the traditional 'input' / 'output' indicators, a new category indicators that has been emerging since the late nineties tries to describe the 'process' of knowledge creation and diffusion within the systems of innovation. Such indicators examine for instance the existence of networks of researchers/inventors, the extent to which the industrial base makes use of the results of the scientific work for its innovative activities etc.

A lot of work is going on in this field, particularly on :

- Co-publishing activities;
- Co-patenting activities;
- Patents-to-patents citations/references;

- Patents-to-'Non-Patent-Literature' (e.g. scientific publications) citations/references;
- Concordance tables linking NACE sectors and IPC patent codes;
- Concordance tables linking NACE sectors (NACE Rev1), OECD Fields of Science (FoS) and UNESCO Fields of education (ISCED1997), with the aim to provide a better insight at the science base of high-tech sectors;

Although most of these studies provide new and extremely valuable insights, there is still a lack of co-ordination and harmonisation between the methods and approaches taken, resulting in a lack of comparability across time and space. More work needs to be done to better integrate the experimental studies in the official Statistical systems.

### B. Internationalisation of R&D and flows of R&D investment

As regards international flows of R&D spending, the major gap in our information refers to the R&D expenditure made by EU Companies outside their territory, and particularly outside the Triad. The new European FATS regulation (Foreign Affiliates Statistics -which should be adopted at the beginning of 2006) includes the collection of R&D expenditure data for 'Inward R&D' (R&D spending in the EU by affiliates of foreign companies), but does not foresee any data collection on 'Outward R&D' (R&D expenditure by EU Member States abroad).

## *(iii) Beyond national aggregates*

### A. Sector-specific R&D expenditure data

The interest for -and use of- sector-specific R&D policies and R&D specialisation is hampered by the fact that the current Nace classification is not very appropriate for measuring the knowledge economy and society. In particular:

- There is a lack of national and international consistency in attributing the NACE codes (e.g. in classification of 'service companies' with regard to the EU and the US);
- The NACE codes should be listed on different levels of aggregation to serve the purpose of R&D specialisations (e.g. pharmacy as sub-category besides the rest of chemistry);
- The allocation of sectoral codes to diversify the R&D budgets of multi-activity R&D is not supported by methodological guidelines.

With respect to technology-specific information, of relevance for instance for analysis of biotechnology, nanotechnology, new materials, ICT… , there is not yet an internationally standardized and commonly used classification and reporting by technology area.

The sectoral activity classification Nace/Isic is currently under revision, in creating the Nace version 2. This improved activity classification takes much better care of the needs of measuring the knowledge economy and society, in rebalancing the classification towards services.

DG ESTAT and the OECD plan to implement the new Nace classification for R&D statistics from the reference year 2008 onwards. This should considerably improve the usefulness of the sectoral R&D data.

## B. Regional statistics

The current Eurostat legislation requires Member States to provide some limited R&D data at regional (NUTS2) level. However, this is insufficient for making detailed analyses of innovation hotbeds such as Universities and Science Parks. European S&T systems are in a stage of profound change, being re-shaped, among other phenomena, by the emergence of new actors becoming central to their dynamics and performance and the European Research Area (ERA) as a full dimension of policy.

## C. 'Micro'-indicators (institutions / organisations level)

Over the recent years, several new 'indicators' and statistical bases have been developed that describe/measure core actors of the S&T systems. Openings in this direction could be seen in the collection of more micro-data and in the micro-data linking between several data sources.

With respect to firms, several surveys & questionnaires, (OECD R&D survey, Community Innovation Survey, R&D Scoreboard (using company account information), are organized using the individual firm's unit of analysis. When firm identification would be more standardized, information could be pooled across surveys and other databases (eg patents, value added, trade…).

With respect to other non-profit agents, like universities, research centers, the OECD collects information on R&D budgets and personnel at the level of the these individual organisations. But this non-profit survey is less internationally standardized, as compared to the profit sector.

Beyond the "official" surveys, there are also a number of more "ad hoc" surveys on individual units. (Eg surveys on TTOs,…). Most of these exercises, however, have been developed with a lack of co-ordination at the international level. As a consequence, most of them cannot (yet) provide enough comparability across countries and periods (no panel structure). There is a need here for more involvement from the European Statistical System/Eurostat and the OECD in this field.

# Conclusions

There is clearly loads of work to do to improve S&T relevant statistics.

1. improving the quality of existing data and indicators;
2. developing new series of indicators for new areas;
3. developing more disaggregated sets of indicators,
4. improving the linking of data & indicators.

But beyond the "creation" of better statistics, it is important to improve on the "diffusion" of S&T statistics. Data should be more easily accessible by the relevant users: policy makers but also researchers, who serve as an important intermediary to process the information into policy analysis. Giving researchers more easy access to data is therefore an essential component of getting more evidence based policies.

Finally, the process of creation and diffusion of S&T statistics, should be less linear, more interacted. Users/researchers should be more actively involved in the process of design of the S&T statistics, so that they can inform the Statistical System of user needs, but also better understand the technical constraints of the Statistical System.

# Appendix: Enterprise Innovation Scoreboard (EIS) Indicators

***1. Human resources***

1.1 S&E graduates (‰ of 20 - 29 years age class)
1.2 Population with tertiary education (% of 25 - 64 years age class)
1.3 Participation in life-long learning (% of 25 - 64 years age class)
1.4 Employment in medium-high and high-tech manufacturing (% of total workforce)
1.5 Employment in high-tech services (% of total workforce)

***2. Knowledge creation***

2.1 Public R&D expenditures (GERD - BERD) (% of GDP)
2.2 Business expenditures on R&D (BERD) (% of GDP)
2.3.1 EPO high-tech patent applications (per million population)
2.3.2 USPTO high-tech patent applications (per million population)
2.4.1 EPO patent applications (per million population)
2.4.2 USPTO patents granted (per million population)

***3. Transmission and application of knowledge***

3.1 SMEs innovating in-house (% of manufacturing SMEs and % of services SMEs)
3.2 SMEs involved in innovation co-operation (% of manuf. and services SMEs)
3.3 Innovation expenditures (% of all turnover in manufacturing and services)

***4. Innovation finance, output and markets***

4.1 Share of high-tech venture capital investment
4.2 Share of early stage venture capital in GDP
4.3.1 SMEs sales of 'new to market' products (% of all turnover in manufacturing and services SMEs)
4.3.2 SME sales of 'new to the firm but not new to the market' products (% of all turnover in manufacturing and services SMEs)
4.4 Internet access/use
4.5 ICT expenditures (% of GDP)
4.6 Share of manufacturing value-added in high-tech sectors
4.7 Volatility-rates of SMEs (% of manufacturing and services SMEs)

# Appendix: RTD Indicators

*Investment*

- Share of gross domestic expenditures on R&D (GERD) in GDP
- GERD as a % of GDP by source of fund
- Share of business enterprise expenditures on R&D (BERD) in GERD
- Share of BERD financed by government
- Share of SMEs in BERD financed by government
- R&D intensity (R&D expenditures as % of GDP) across industries in manufacturing
- R&D intensity (R&D expenditures as % of GDP) in some high-tech sectors

*Human Resources*

- Share of total tertiary education expenditure in GDP
- Share of researchers (RSEs) in population
- Share of R&D personnel in labour force by institutional sector
- R&D expenditures by RSE by institutional sector
- Number of yearly new S&T PhD in 25-34 population
- Breakdown of employed HRST according to native country

*Innovation potential*

- Number of patents with EPO and USPTO
- Number of High-Tech patents at EPO and USPTO per capita

*Business Innovation*

- Share of seed& start-up venture capital in GDP
- Share of seed& start-up in venture capital for all sectors and for high-tech sectors
- Expenditure on innovation in turnover of manufacturing industry
- SMEs innovating inhouse (% of manufacturing SMEs)
- Innovative co-operating SMES (% of manufacturing SMEs)

*Competitiveness*

- Technology Balance of Payments per capita
- High-tech products imports and exports per capita

(Source: DG RESEARCH, Investing in research: an action plan for Europe).