Statistics and Science, Technology and Innovation Policy:
How to Get Relevant Indicators

Benoît Godin

385, rue Sherbrooke Est
Montreal, Quebec
Canada H2X 1E3
benoit.godin@ucs.inrs.ca

OECD Blue Sky II Conference
"What Indicators for Science, Technology and Innovation Policies
in the 21st Century"
Ottawa, Canada
25-27 September 2006

Statistics and Science, Technology and Innovation Policy: How to Get Relevant Indicators

I would like to contribute to this conference with some perspectives, namely historical perspectives concerning statistics on science, technology and innovation, and use such a perspective to suggest some avenues to develop an agenda for the future. Essentially, I suggest going back to three basic conditions as necessary steps before constructing a new generation of indicators. First, that we start to reconsider critically most of the conceptual frameworks that we now use to collect and analyze statistics. Secondly, that we start thinking seriously about what national systems really mean for statistics, instead of focusing on international comparisons and standardization of methodologies. Third, that we depart for a moment from the economic approach. These ideas are extreme and provocative, but I am sure all of you can adapt them to your own take on the situation.

Most of us present today probably do not know that 2006 marks another anniversary besides that of the Blue Sky Forums (the first Forum having been held ten years ago). This anniversary is the centennial of statistics on science. [1] In 1906, James McKeen Cattell, an American psychologist and the editor of *Science* from 1895 to 1944, launched a biographical directory on American scientists. From the information contained in the directory, Cattell published systematic quantitative studies on science. For over thirty years, he measured the number of scientists, their demography, their geography and what he called their performance. [2] He would soon be followed by his peers from the discipline of psychology, who pioneered the systematic use of counting papers in order to measure scientists' production of knowledge. [3]

Measuring the number of scientists rather than other aspects on science had to do with the context of the time. Cattell was a student of the British statistician Francis Galton, who proposed, in the second half of the 19th Century, (positive) eugenics as a solution to the population problem. To many people at the time, the stock of the population and the

[1] See http://www.csiic.ca/index.html.
[2] B. Godin (2007), From Eugenics to Scientometrics: Galton, Cattell and Men of Science, *Social Studies of Science*, forthcoming.
[3] B. Godin (2006), On the Origins of Bibliometrics, *Scientometrics*, 68 (1): 109-133.

quality of the race was deteriorating, and those groups that contributed more to civilization, namely eminent men and including scientists, were not reproducing enough. The "unfits" were far more productive, and some suggested policies for sterilizing them. Hence the idea of measuring the number of available scientists, the size of the scientific community and the contribution of scientists to the "races". Obviously, no scientist thought about sterilizing those scientists who published bad papers. This would have solved many problems currently plaguing science and the scientific journal system. The trouble is that my dean may come down and sterilize me. He is fully capable of this, and can do it with a mere look. [4]

After World War I, and increasingly so after World War II, a completely new type of statistics appeared. In fact, by that time it was no longer scientists like Galton or Cattell who produced statistics on science, but governments and their statistical bureaus. And it was no longer the number of university scientists the bureaus were interested in, but the money spent on research. [5] This had to do, again, with the context of the time: the cult of efficiency and the performance of organizations. Research was considered as the vehicle toward economic prosperity, and organizations and their "organized" laboratories were seen as the main vector to this end. To statisticians and policy analysts, the "research budget", or Gross Expenditures on Research and Development (GERD), as the sum of expenditures in four groups of organizations, or economic sectors, became the most cherished indicator.

The main consequence of such an orientation for statistics was twofold. First, statistics came to be framed into an accounting framework. Statistics on science, technology and innovation concentrated on costs, aligning themselves with the System of National Accounts, and collected within an input/output model. Most current indicators are economic in type: expenditures on research, output such as patents, high-technology products, marketed innovation, etc. You would look in vain for systematic indicators on the social side of science. The second consequence was a focus on productivity. Certainly, the concept of productivity in science arose from scientists themselves. In Galton's hands, productivity meant reproduction: the number of children a scientist had,

---

[4] I owe this "inappropriate" joke to Stephen J. Bensman, Louisiana State University (Los Angeles). Personal conversations, 18 and 21 May 2006.
[5] B. Godin (2005), *Measurement and Statistics in Science and Technology : 1920 to the Present*, London : Routledge.

or the number of scientists a nation produces. Then, in the 20th Century, productivity came to mean the quantity of output of a scientific or technological type, and later economic (labour or multifactor) productivity, or outcomes of research on economic growth. Today, It is the organizations (and the economic sector to which they belong) that is measured and examined, above all firms (think of the innovation surveys), and not the individuals or groups who compose them, nor the people from society who are supposed to benefit from science.

There are at least three reasons that explain this orientation in current statistics. One is the basic unit of science policy and analysis. Whereas early studies of science, technology, and innovation particularly sociological studies, were concerned with people and the varied impacts of science on people's lives, current studies focus entirely on efficiency. Economic growth, productivity, and profitability rather than quality of life, drives policies. Second, and methodologically, economic output is easier to measure than, for example, the social and cultural aspects or impacts of science. For this reason, many researchers use data sources that are easily available and standardized rather than developing specific surveys. Third, most studies are conducted by economists or, for purposes of "emulation", by researchers using an economic-type framework. These, then, are three factors that automatically suggest three *loci* for improving statistics: the policy frameworks, the sources of data, and … the researchers.

Let me make three suggestions for an agenda for the future. First, we need to abandon entirely the current policy frameworks. Actually, the field of science and innovation studies, particularly the policy-oriented and the statistical subfields, has fully endorsed the productivity issue. Every conceptual framework developed over the last fifty years is concerned with firms and with accounting and efficiency in the broadest sense. Whether you look at models of innovation, evaluation exercises such as input-output analysis, policy frameworks on the information economy or society, the national system of innovation, the knowledge-based economy, or the new economy, the most central issue and the statistics are economic, among which is the concept of productivity (with comparisons with the United States as ideal). Either one measures the productivity of the science system itself, or scientific productivity (academic papers), or the contribution of science to economic growth and productivity. We need to look at science anew, and forget, for a moment, economic issues. Statistics should look more systematically at how science contributes to social issues and policies.

To this end, and this is my second suggestion, we need completely new data sources. I am not particularly enthusiastic about using existing data for new indicators. What we will get will be more of the same. What we need here is, first, to move from economic datasets to multidimensional measures of science: social, cultural, health, environment, etc. [6] Measuring the impacts of science, technology and innovation, for example, means looking at **changes** on nature, society and people such as changes on understandings, beliefs (and attitudes), and behaviours. Admittedly, the challenges are many for anyone concerned with measuring intangible outcomes of a social type. But weren't measurements of science as challenging in the 1950s and 1960s when governments started collecting statistics on research expenditures? There are currently several initiatives in many countries looking precisely at how to measure outcomes other than strictly economic outcomes. Unfortunately, these initiatives are not conducted by statistical offices, but by government departments, for their own *ad hoc* needs and not necessarily for developing indicators of a systematic nature.

The other urgent task as regards data sources is to move from macro and aggregate statistics to more detailed ones. Currently, users of statistics are asking for a lot more information than before because their analyses and/or decisions are more fine-grained. National aggregates are no longer enough, and neither are standard classifications. This means that we should not seek international standardization for its own sake. Too often, statisticians' efforts are religiously devoted to adapting to international frameworks and methodological norms. Rather we need data that are adjusted to the national and local situations we want to measure. Maybe it is time to rethink international comparisons as the ultimate objective of statistics.

My last suggestion concerns the producers of statistics. As a rule, and as decades of sociological studies have shown, you have to change people to get new ideas. I therefore suggest sterilizing economists. Not because they produce bad things, but

[6] Godin and Doré have identified eleven dimensions for measuring the impacts, or outcomes of science: knowledge, training, technology, economy, culture, society, policy, organizations, health, environment, symbolic. B. Godin and C. Doré (2005), *Measuring the Impacts of Science: Beyond the Economic Dimension*, INRS: Montréal.

because they produce too much, eclipsing others' work. If we really want new types of statistics, we should try looking for ideas from a more diverse range of disciplines and approaches. Since I do not belong to any discipline – I am a multidisciplinary, or rather undisciplined researcher – I believe I am not really in conflict of interest with this recommendation.