The Culture of Numbers:
The Origins and Development of Statistics on Science

Benoît Godin

INRS

385 rue Sherbrooke Est
Montreal (QUE)
Canada H2X 1E3
benoit.godin@ucs.inrs.ca

Project on the History and Sociology of STI Statistics

Working Paper No. 40

2009

Previous Papers in the Series:

1. B. Godin, *Outlines for a History of Science Measurement*.
2. B. Godin, *The Measure of Science and the Construction of a Statistical Territory: The Case of the National Capital Region (NCR)*.
3. B. Godin, *Measuring Science: Is There Basic Research Without Statistics?*
4. B. Godin, *Neglected Scientific Activities: The (Non) Measurement of Related Scientific Activities*.
5. H. Stead, *The Development of S&T Statistics in Canada: An Informal Account*.
6. B. Godin, *The Disappearance of Statistics on Basic Research in Canada: A Note*.
7. B. Godin, *Defining R&D: Is Research Always Systematic?*
8. B. Godin, *The Emergence of Science and Technology Indicators: Why Did Governments Supplement Statistics With Indicators?*
9. B. Godin, *The Number Makers: A Short History of Official Science and Technology Statistics*.
10. B. Godin, *Metadata: How Footnotes Make for Doubtful Numbers*.
11. B. Godin, *Innovation and Tradition: The Historical Contingency of R&D Statistical Classifications*.
12. B. Godin, *Taking Demand Seriously: OECD and the Role of Users in Science and Technology Statistics*.
13. B. Godin, *What's So Difficult About International Statistics? UNESCO and the Measurement of Scientific and Technological Activities*.
14. B. Godin, *Measuring Output: When Economics Drives Science and Technology Measurements*.
15. B. Godin, *Highly Qualified Personnel: Should We Really Believe in Shortages*?
16. B. Godin, *The Rise of Innovation Surveys: Measuring a Fuzzy Concept*.
17. K. Smith, *Measurement of Innovation in Europe: Concepts, Experience and Results.*
18. B. Godin, *A Note on the Survey as Instrument for Measuring S&T.*
19. B. Godin, *Rhetorical Numbers*: *How the OECD Constructs Discourses on Science and Technology.*
20. B. Godin, *Are Statistics Really Useful? Myths and Politics of Science and Technology Indicators.*
21. B. Godin, *The New Economy: What the Concepts Owes to the OECD.*
22. B. Godin, *The Most Cherished Indicator: Gross Domestic Expenditures on R&D (GERD).*
23. B. Godin, *Technological Gaps: Quantitative Evidence and Qualitative Arguments.*
24. B. Godin, *The Knowledge-Based Economy: Conceptual Framework or Buzzword*?
25. B. Godin, *The Obsession for Competitiveness and its Impact on Statistics: The Construction of High Technology Indicators*.
26. B. Godin, *Globalizing Indicators: How Statisticians Responded to the Political Agenda on Globalization*.
27. B. Godin, *The Who, What, Why and How of S&T Measurement*.
28. B. Godin, *Research and Development: How the "D" Got Into R&D*.
29. B. Godin, *Technological Progressiveness as a Precursor to the Concept of High-Technology*.
30. B. Godin, *The Linear Model of Innovation: the Historical Construction of an Analytical Framework.*
31. B. Godin, *Science, Accounting and Statistics: the Input-Output Framework and its Role as Metaphor.*
32. B. Godin, *From Eugenics to Scientometrics: Galton, Cattell and Men of Science*.
33. B. Godin, *On the Origins of Bibliometrics*.
34. B. Godin, *The Value of Science: Changing Conceptions of Productivity, 1869-circa 1970.*
35. B. Godin, *What is Science: Defining Science by the Numbers, 1920-2000*.
36. B. Godin, *National Innovation System: The System Approach in Historical Perspective*.
37. B. Godin, *The Knowledge Economy: Fritz Machlup's Construction of a Synthetic Concept*
38. B. Godin, *The Information Economy: the History of a Concept Through its Measurement, 1949-2005*.
39. B. Godin, *The Making of Statistical Standards: The OECD and the Frascati Manual, 1962-2002*.

Project on the History and Sociology of STI Statistics
385 rue Sherbrooke Est, Montreal, H2X 1E3
Telephone: (514) 499-4074 Facsimile: (514) 499-4065
www.csiic.ca

Abstract

Measuring science has become an ‘industry’. When, how and why did science come to be measured in the first place? How did a “cultural” activity – science – long reputed to be not amenable to statistics, come to be measured? The statistics owes its existence to the context of the time: 1) measuring the contribution of great men, among them scientists, to civilization, and improving the social conditions of scientists; then 2) science policy and the efficiency of investments in research.

Before the 1920s, it was scientists themselves who conducted measurements on science. The statistics collected concerned men of science, or scientists, their demography and geography, their productivity and performance, and were used to promote what was called the advancement of science. In the 1940s and after, the kind of statistics collected changed completely. It was no longer scientists who collected them, but governments and their statistical bureaus. The most cherished statistics was thereafter money devoted to research and development (R&D).

The Culture of Numbers:
The Origins and Development of Statistics on Science

Measuring science has become an "industry". Governments and their statistical offices have conducted regular surveys of resources devoted to research and development (R&D) since the 1960s. The methodology used is that suggested and conventionalized by the OECD Frascati manual, adopted by member countries in 1963, and now in its sixth edition. [1] Since the 1990s, national governments have also conducted regular surveys on innovation, again based on an OECD methodology known as the Oslo manual. [2] More recently, scoreboards of indicators have appeared that collect multiple indicators on science, technology and innovation (thereafter science). [3]

The statistics collected by official organizations are regularly used by academics, among them economists who, over the last five decades, have produced a voluminous literature on measuring the contribution of science to economic growth and productivity. [4] Academics are also producers of their own statistics. Using scientific paper-counts as a tool, sociologists and others have studied the "productivity" of scientists since the early 1900s. [5] Today, a whole community of researchers concerned with counting papers and citations call themselves bibliometricians.

When, how and why did science come to be measured in the first place? How did a "cultural" activity – science – long reputed to be not amenable to statistics, come to be measured? This paper is dedicated to tracing the origin and development of statistics on science, and the impact statistics have had on the representation of "science" over the 20th

[1] OECD (1962), *The Measurement of Scientific and Technical Activities: Proposed Standard Practice for Surveys of Research and Development*, Paris.
[2] OECD/Eurostat (1997), *Proposed Guidelines for Collecting and Interpreting Technological Innovation Data* (Oslo Manual), Paris.
[3] The OECD has published a biennial publication titled *Science, Technology and Industry Scoreboard* since 1995, and the European Commission has published an *Innovation Scoreboard* since 2001.
[4] Z. Griliches (1998), *R&D and Productivity: The Econometric Evidence*, Chicago: University of Chicago Press.
[5] B. Godin (2006), On the Origins of Bibliometrics, *Scientometrics*, 68 (1), pp. 109-133.

century. It documents two stages in this history. Before the 1920s, it was scientists themselves who conducted measurements on science. The statistics collected concerned men of science, or scientists: their demography and geography, their productivity and performance. This kind of statistics owes its development to the context of the time: measuring the contribution of great men, among them scientists, to civilization; then, improving the social conditions of scientists.

Starting in the 1940s, the kind of statistics collected changed completely. It was no longer scientists who collected them, but governments and their statistical bureaus. The most cherished indicator was thereafter money devoted to research and development (R&D). Again, this owes its development to the context of the time, namely science policy and efficiency. Science policy developed primarily due to concerns about using accounting as a way of controlling (government) expenses on R&D. But second, official statistics also developed for a more positive aim: to determine target levels for the investment in scientific activities for public goods.

The first part of this paper documents how the context of eugenics in the second half of the nineteenth century, namely the will to improve the quality of the populations, led to counting 'men of science' as part of that class of great men responsible for the progress of civilization. The second part shows how, as the context changed and the policy issues shifted to economic progress over the twentieth century, a new king of statistics developed: accounting. The last part analyzes the impact of statistics on representations of science.

Eugenics, Men of Science and Productivity

The measurement of science emerged out of interest in great men, heredity and eugenics, and the contribution of eminent men to civilization. Among these eminent men were men of science, the population of whom was thought to be in decline and insufficiently appreciated and supported. Statistics thus came to be collected to document the case, and to contribute to the advancement of science – and of the scientific profession. The statistics conceived were concerned with measuring the size of the scientific community, or men of science, and their social conditions. [6]

British statistician Francis Galton's (1822-1911) measurements of science, the first to be conducted worldwide, were specifically based on his belief that the progress of civilization depends on great men, whose numbers were in decline. Enunciating these views, Galton suggested: "the qualities needed in civilized society are, speaking generally, such as will enable a race to supply a large contingent to the various groups of eminent men". [7] To Galton, however, there were only 233 eminent British men for every one million population, while "if we could raise the average standard of our race one grade" there would be 2,423 of them.

Thus, Galton elected to pursue the notion of genius. *Hereditary Genius*, published in 1869, had two purposes: measuring intellectual ability in a population, and documenting the role of heredity in the transmission of intellectual ability. Among other things, he calculated that men of science were exceptionally productive of eminent sons, and this he attributed to family environment (other professional groups attributed it to heredity).

Five years after *Hereditary Genius*, Galton turned his attention entirely to this one specific group of illustrious men – men of science. In *English Men of Science*, [8] Galton drew up a list of 180 men – out of 300 existing British men of science as he estimated, or 1 in every 10,000 population. Analysis of their antecedents revealed that men of science had less children than their parents had: the number of their living children between the ages of 5 and 50 was on average 4.7, compared to 6.3 in the families these men of science came from. To Galton, the numbers revealed a clear "tendency to an extinction of the families of men who work hard with the brain", "a danger to the continuance of the race".

Galton concentrated on men of science again in 1906 for the third and the last time in his life. *Noteworthy Families* was "to serve as an index to the achievements of those families

[6] B. Godin (2007), From Eugenics to Scientometrics: Galton, Cattell and Men of Science, *Social Studies of Science*, 37 (5), pp. 691-728.
[7] F. Galton (1869), *Hereditary Genius: an Inquiry into Its Laws and Consequences*, Honolulu, University Press of the Pacific, 2001.
[8] F. Galton (1874), *English Men of Science: Their Nature and Nurture*, London: Macmillan.

which [have] been exceptionally productive of noteworthy persons”. [9] Galton sent a questionnaire to all living fellows of the Royal Society in the spring of 1904. He also drew names from biographical dictionaries. In total, he sent out 467 questionnaires and received 207 replies. He retained 100 completed returns for statistical purposes, corresponding to 66 families. Galton found again that “a considerable proportion of the noteworthy members in a population spring from comparatively few families”. He estimated this proportion of noteworthy persons to the whole population as 1 to 100. The main result of his study, however, was a reduction in the estimated population of noteworthy men. Galton observed 207 noteworthy members in the families studied, compared to a statistical expectation of 337.

Galton’s works on men of science have been very influential. *English Men of Science* was the first quantitative “natural history” or “sociology” of science, as he himself called it. [10] *English Men of Science* relied on a dedicated survey among a specific group of men, while most studies of eminent men were based on statistics constructed from biographical dictionaries, as *Hereditary Genius* had been, or on institutional data, like membership in scientific societies. Certainly, in the mid-1850s, censuses began collecting information on professions, among them teachers and professors, and could have been used to measure science. But the category “men of science” (or scientists) did not exist in the classifications used. Galton must be credited with having offered the first quantitative estimates regarding the number of men of science in a population. He would soon be followed by others.

In 1895, the US psychologist James McKeen Cattell (1860-1944) acquired the weekly journal *Science*, established in 1880 by Alexander Graham Bell and Gardiner G. Hubbard. A few years after acquiring the journal, Cattell’s research on mental testing was judged fuitless. He had initiated a large-scale program of testing Columbia students every year, similar to Galton’s experiment in museums and public expositions. In the end,

[9] F. Galton and E. Schuster (1906), *Noteworthy Families (Modern Science): An Index to Kinships in Near Degrees between Persons Whose Achievements Are Honourable, and Have Been Publicly Recorded*, London: John Murray.

[10] Published at the same time as one by the biologist Alphonse de Candolle. See A. de Candolle (1873), *Histoire des sciences et des savants depuis deux siècles, d’après l’opinion des principales académies ou sociétés scientifiques*, Paris: Fayard, 1987.

however, it appeared that he was measuring psychological behaviour (like alertness) rather than mental abilities, and he was criticized for this. Cattell partly redirected his efforts away from experimental psychology. Besides editing *Science* and other journals, Cattell, as a student of Galton, turned to another kind of statistical analysis than experimental psychology: the "scientific" study of science. To Cattell, applying statistics to study men of intelligence, above all men of science, was highly desirable: "the accounts of great men in biographies and histories belong to literature rather than to science (…). It is now time that great men should be studied (…) by the methods of exact and statistical science". [11] There was a specific motive behind such studies, a motive learned from Galton. In an early study on eminent men, Cattell asked: "Are great men, as Carlyle maintains, divinely inspired leaders, or are they, as Spencer tells us, necessary products of given physical and social conditions? (…). We can only answer such questions by an actual study of facts". And he continued as follows: "We have many books and articles on great men, their genius, their heredity, their insanity, their precocity, their versatility and the like, but, whether these are collections of anecdotes such as Professor Lombroso's or scientific investigations such as Dr. Galton's, they are lacking in exact and quantitative deductions (…). Science asks how much? We can only answer when we have an objective series of observations, sufficient to eliminate chance errors (…)". Cattell's concrete proposal was to observe, classify, measure and compare.

As a first step in this program, Cattell selected 1,000 men from six biographical dictionaries or encyclopedias in order to study the distribution of eminence among nations. The statistics showed that only a few nations produce eminence: "France leads, followed pretty closely by Great Britain. Then there is a considerable fall to Germany and Italy". To Cattell, the moral was clear: "The progress to our present civilization may have depended largely on the comparatively few men who have guided it, and the civilization we hope to have may depend on a few men (…). If we can improve the stock by eliminating the unfit or by favoring the endowed – if we give to those who have and take away from those who have not even that which they have – we can greatly accelerate and direct the course of evolution. If the total population, especially of the well endowed, is

[11] J. M. Cattell (1903), A Statistical Study of Eminent Men, *Popular Science Monthly*, February, pp. 359-377.

larger, we increase the number of great men". As a continuation of this study, Cattell devoted his efforts to men of science. However, he soon changed his mind on heredity, and argued for the improvement of the social conditions of men of science.

Between 1902 and 1906, Cattell constructed a directory (called *American Men of Science*) for a contract granted by the newly-created Carnegie Institution of Washington (1902). As Cattell recalled, "Mr. Carnegie has specified as one of the main objects of his foundation, to discover the exceptional man in every department of study whenever and wherever found, inside or outside of schools, and enable him to make the work for which he seems specially designed his life work". But how to find exceptional men? How to distribute money among fields?

Compiling a biographical directory was Cattell's suggestion. The first edition contained about 4,000 biographical sketches of men of science, restricted to those men "who have carried on research work" and "contributed to the advancement of pure science" (natural science). By 1944, the last year Cattell edited the directory before he died, the document contained biographical information on over 34,000 men of science. From the directory, Cattell constructed statistics.

Two concepts were fundamental to his work. The first was **productivity**, defined as the number of men of science a nation produces. Cattell compared American states and institutions in terms of both absolute and relative (per million population) numbers of men of science. He found concentrations of origin in a few regions: Massachusetts and Boston were identified as the intellectual center of the country, while the South "remains in its lamentable condition of scientific stagnation". To Cattell, this fact contradicted Galton's thesis: "the inequality in the production of scientific men in different parts of the country seems to be a forcible argument against the view of Dr. Galton and Professor Pearson that scientific performance is almost exclusively due to heredity. It is unlikely that there are such differences in family stocks as would lead one part of the country to produce a hundred times as many scientific men as other parts (…). The main factors in producing scientific and other forms of intellectual performance seem to be density of population, wealth, opportunity, institutions and social traditions and ideals". According to Cattell, "the scientific productivity of the nation can be increased in quantity, though not in quality, almost to the extent that we wish to increase it". [12]

To Cattell, "eminent men are lacking and this we must attribute to changes in the social environment": the growing complexity of science, educational methods, lack of fellowships and assistantships as well as prizes, teaching load, and low salary. "The salaries and rewards are not adjusted to performance", unlike Germany, Great Britain and France, where the "exceptional men have received exceptional honors (…). Methods should be devised by which scientific work will be rewarded in some direct proportion to its value to society - and this not in the interest of the investigator, but in the interest of society".

Productivity was the first concept Cattell introduced in his statistical analyses. The second was that of **performance**. Whereas productivity measured quantity, performance measured quality or merit, defined as "contributions to the advancement of science, primarily by research". Cattell believed that "expert judgment is the best, and in the last resort the only, criterion of performance". [13] He thus asked ten leading representatives of each of the twelve sciences he selected to arrange the men of science, whose names appeared in the directory, in order of merit (rank). The "positions assigned to each man were averaged, and the average deviations [probable error] of the judgments were calculated [and individuals arranged in order]".

Cattell compared his procedure of votes to that used in elections to a scientific society, or in filling chairs at a university. His method was said to be superior: "the academy has no method of comparing performance in different sciences". [14] To Cattell, "the methods of selection used in this research are more accurate than those of any academy of sciences, and it might seem that the publication of the list would be as legitimate as that of a list of

[12] J. M. Cattell (1906), A Statistical Study of American Men of Science III: The Distribution of American Men of Science, *Science*, 24 (623), December 7, pp. 732-742.
[13] J. M. Cattell (1906), A Statistical Study of American Men of Science II: The Measurement of Scientific Merit, *Science*, 24 (622), November 30, pp. 699-707.
[14] J. M. Cattell (1906), A Statistical Study of American Men of Science: The Selection of a Group of One Thousand Scientific Men, *Science*, 24 (621), November 23, pp. 658-665.

our most eminent men selected by less adequate methods. But perhaps its very accuracy would give it a certain brutality".

What Cattell observed from the distribution of the top-ranked (or "starred", i.e.: marked with an asterisk on the list) one thousand scientists would become a fact much studied later in the literature – that the distribution of merit follows an "exponential law" rather than the normal distribution of ability shown in Galton's work. Measuring performance allowed Cattell to estimate gains and losses in ranks or places: those men of science who have attained a place in the one thousand and those who have lost their place over time. Cattell then ranked institutions by the order of merit of their scientific men, and offered his readers the first league table of universities in the history of statistics on science. "I give this table with some hesitation, but it appears in the end it will be for the advantage of scientific research if it is known which institutions obtain and retain the best men (...). A table such as this might have some practical influence if the data were made public at intervals of ten years". [15] The table showed Harvard, Columbia and Chicago as the leading universities in terms of their share of the top thousand scientific men. All in all, Cattell calculated that about half of the top thousand scientific men were connected with just 18 institutions.

Cattell would continue analyzing statistics on men of science on this same line up until the 1930s, looking at changes that took place in the distribution of sciences, and in the origins and position of scientific men since the last series of data. Cattell also made use of some data on publications (a specialty now called bibliometrics) to measure the progress of science. The systematic use of bibliometrics, however, was pioneered by other American psychologists. One specific aim guided their efforts: to defend the status of psychology as a science.

Statistics on psychological science were specifically developed to contribute to the advancement of the discipline of psychology. [16] Using paper counts, psychologists showed with confidence how psychology was really a science among the sciences. While the yardstick for comparing the scientific profession in America was Europe, reputed for

---

[15] J. M. Cattell (1906), A Statistical Study of American Men of Science III, *op. cit.*
[16] B. Godin (2006), On the Origins of Bibliometrics, Scientometrics, *op. cit.*

its chairs, laboratories and public support, for the science of psychology it was its status *vis-à-vis* the other sciences, experimental in character, that served as the benchmark. To Cattell, for example, "compared with psychology, a science such as astronomy may almost be regarded as naïve. The entire known performance of the solar system and of the fixed stars since the time of the Chaldaean is less complicated than the play of a child in its nursery for a single day (…). Atoms and molecules are so invisible, the ether is so intangible, we know after all so little about them, that it is easy to invent hypotheses". And he continued: "The two greatest scientific generalizations of the present century are the conservation of energy and evolution by survival of the fit. Now, if consciousness alters, however slightly, the position of molecules in the brain the fundamental concept of physical science must be abandoned. If consciousness has no concern in the actions of the individual we have one of the most complex results of evolution developed apart from the survival of useful variations, and the Darwinian theory has failed (…). The world is one world; every part of it is in relation to every other part, and each part consists in these relations". [17]

Several psychologists developed a rhetoric on progress in psychology ("**taking stock of progress**", as psychologist E. F. Buchner called it) [18] in which measures of growth were constructed for psychologists (their absolute number, geographical distribution, number per million population, status, degrees), curricula, doctorates conferred, laboratories, journals and … publications. Two vehicles carried these numbers. The first was periodic reviews. The second vehicle for assessing the progress made in psychology was histories of the Association.

It was S. W. Fernberger of the University of Pennsylvania who would further develop the statistics on publications. Fernberger is well known today for having produced "classics" in the history of psychology. [19] With regard to papers, he noticed the increasing emphasis placed on publishing as a criterion for eligibility in psychological Associations. He

[17] J. M. Cattell (1898), The Advance of Psychology, *Science*, 8 (199), October 21, pp. 533-54.
[18] E. F. Buchner (1903), A Quarter Century of Psychology in America, 1878-1903, *American Journal of Psychology*, July-October, p. 57.
[19] S. W. Fernberger (1932), The American Psychological Association: a Historical Summary, 1892-1930, *Psychological Bulletin*, 29 (1), pp. 1-89; S. W. Fernberger (1943), The American Psychological Association: a Historical Summary, 1892-1942, *Psychological Review*, 50 (3), pp. 33-60.

charted the number of papers presented at each meeting since 1892, and looked at the "productivity" of universities at these meetings, measuring that 19 universities produced 53% of all papers. Then, in 1917, he started a series of studies on the scientific production of nations entitled *National Trends in Psychology*. These were published at intervals of ten years from 1917 to 1956. [20] Fernberger documented German supremacy in the first decades of the twentieth century, then a decline; English titles were shown to be on an upward trend, while French titles declined.

Equally noteworthy is S. I. Franz's paper of 1917. Professor at George Washington University (1906-1921), Franz produced a bibliometric study on the scientific performance of psychologists. [21] "Within the past few years there have appeared reviews of the progress of psychology for different periods of time (…)", stated Franz. But "we have not been informed by whom the psychological advances have been made, or whether or not in view of the increasing number of professional psychologists there has been a corresponding increase in the number or in the value of the published investigations. In other words, although it is admitted that advance has been made, we are as far from knowing **whether or not the advance has been satisfactory** and corresponds with the number of psychologists".

To Franz, methods for estimating the value of individuals' contributions (elections to Academies, selection and promotion in universities) all have defects. "We can do something [more] definite by determining that a certain individual has or has not made any published contribution towards psychological advance". Franz observed a fairly gradual increase in publications over time. But the productivity, now defined as the number of publications by researcher, varied. Franz measured that older men were more productive than younger ones, but the ratio of actual to expected publications was higher among the younger ones. "It seems unlikely that as many as 40% of the older group are engaged in the accumulation of material for the development of a cosmology, or of a system of psychology, or of an exhaustive history of the science, or of other large

[20] S. W. Fernberger, On the Number of Articles of Psychological Interest Published in the Different Languages, *American Journal of Psychology*, 28 (1), 1917, pp. 141-150; 37 (4), 1926, pp. 578-581; 48 (4), 1936, pp. 680-684; 59 (2), 1946, pp. 284-290; 69 (2), 1956, pp. 304-309.
[21] S. I. Franz (1917), The Scientific Productivity of American Professional Psychologists, *Psychological Review*, 24 (3), pp. 197-219.

projects which should not be laid aside in favor of the minor contributions such as articles and monographs (…). The writer feels that some of the so-called "professional" psychologists should be classed with dilettantes". In conclusion "the attention of the reader is called to the consideration of the wisdom of the action of certain scientific societies which require that a member shall retain membership in them only as long as he continues to show an active interest in the advancement of his science by publication".

Statistics on men of science and statistics on scientific papers developed considerably in the following decades. Governments and their statistical bureaus started constructing registers of scientific and technical personnel, then conducted surveys on human resources devoted to research activities. Scientists and their representatives regularly used the data as a rhetorical resource for more public funding of university research. Sociologists and economists, for their part, created a whole "industry, called bibliometrics, and concerned with measuring the output of scientists and studying the factors responsible for scientific productivity.

Accounting of Science

The measurements discussed in the previous section were only the precursors to a long series of statistics produced by governments and their bodies. By the 1940s, it was public organizations that produced most of the statistics and soon got a "monopoly" on the measurement of science, partly because of their financial resources to conduct systematic and regular surveys. It took four years to Cattell to construct its directory on men of science from which he drew statistics. Such investment in time and money are rarely available to individual researchers today. Governments have much more resources.

We owe a large part of the development of **official (or institutional)** measurement of science in western countries to the United States. It was there that the first experiments emerged in the 1920s. Two factors were at work that explained this phenomenon: the need to manage industrial laboratories, and the need to plan government scientific and technological activities, particularly in the event that they might be needed for war (mobilization of scientists). Canada followed a decade later, with the same objectives, and Great Britain in the decade after that. [22]

The very first official measurement of science activities came from the US National Research Council, a body of the National Academy of Sciences. Scientists were thus not only the first statisticians on national scientific activities, [23] but one of their representative organizations was the first to continue their efforts. During World War I, the US National Academy of Sciences convinced the federal government to give scientists a voice in the war effort. The National Research Council was thus created in 1916 as an advisory body to the government. Rapidly, a research information committee, then a Research Information Service, was put into place. The Service was concerned with the inter-allied exchange of scientific information. After the war, however, these activities ceased, and the Service reoriented its work toward other ends. The Service became "a national center of information concerning American research work and research workers, engaged in preparing a series of comprehensive card catalogues of research laboratories in this country, of current investigations, research personnel, sources of research information, scientific and technical societies, and of data in the foreign reports it received". [24] It was as part of these activities that the Service developed directories on research in the United States. Beginning in 1920, the Service regularly compiled four types of directory, the raw data of which were published extensively in the Bulletin of the National Research Council, sometimes accompanied by statistical tables. One directory was concerned with industrial laboratories. The first edition listed approximately 300 laboratories, and contained information on fields of work and research personnel. A second directory dealt with sources of funds available for research, a third with fellowships and scholarships, and a fourth with societies, associations and universities, covering both the United States and Canada. The Council directories were used by many in the following years to conduct statistical analyses of research, particularly industrial research.

[22] B. Godin (2005), *Measurement and Statistics on Science: 1920 to the Present*, London: Routledge.
[23] In this paper, I concern myself with "national" statistics, not with those of public and scientific institutions that may have produced numbers on their own activities, in annual reports for example.
[24] R. C. Cochrane (1978), *The National Academy of Sciences: The First Hundred Years 1863-1963*, Washington: National Academy of Sciences, pp. 240-241.

From the 1940s onward, it was governments that started collecting statistics on money spent on research. These efforts, much influenced by J. D. Bernal in the UK, [25] had forerunners such as the US National Resources Committee, the US President's Research Board, and the US National Science Foundation. [26] What has changed since Cattell is that counting men of science was no longer the statistics *par excellence*. Money devoted to R&D was now the most cherished indicator. Admittedly, Cattell did produce some financial data. Using *Science* as a vehicle, he published several lists of institutional funds (grants) for research starting in 1903, and organized the AAAS Committee of One Hundred, concerned with collecting information on scientific research grants, whose (quite imperfect and incomplete) lists were published between 1916 and 1918. But this kind of data was sporadic.

Two factors explain the new situation. The first was accounting as a way of controlling (government) expenses on R&D, which were, according to bureaus of budget, growing too fast. On a more positive side, and second, statistics were developed on money spent as policy targets for scientific development, and were thus used to convince institutions to devote more money to R&D.

These efforts coalesced into the OECD Frascati manual, written by the British economist C. Freeman. [27] In 1963, the Member countries adopted standards for the measurement of R&D expenditures, and the OECD published a methodological manual. The Frascati manual essentially developed three sets of guidelines. Firstly, norms were proposed for defining science as "systematic" research and distinguishing research from other activities so that the latter could be excluded: these other activities included research/related scientific activities, development/production, and research/teaching. Secondly, the manual suggested classification of research activities according to 1) the sector that finances or executes the research: government, university, industry or non-profit organizations and, in relation to this latter dimension, 2) the type or character of the research, which is either basic, applied or concerned with the development of products

[25] J. D. Bernal (1939), *The Social Function of Science*, Cambridge (Mass.): MIT Press, 1973.
[26] See B. Godin (2005), Measurement and Statistics on Science: 1920 to the Present, *op. cit.*
[27] OECD (1962), The Measurement of Scientific and Technical Activities: Proposed Standard Practice for Surveys of Research and Development, *op. cit*.

and processes, 3) the activities classified by discipline in the case of universities (and non-profit organizations), by industrial sector or product in the case of firms, and by functions or socioeconomic objectives in the case of governments. Finally, the manual suggested a basic statistic as an indicator for policy purposes. [28]

The **GERD** (Gross Expenditures on R&D) is the main statistics which comes out of the manual. It is the total of money spent on R&D by the four following economic sectors: industry, university, government and non-profit. However, the GERD, as it is presented as a statistics on national research or research budget, remains fragile. The first edition of the Frascati manual suggested that national "variations [in R&D statistics] may be gradually reduced" with standardization. But the collection of statistics on R&D expenditures remains a very difficult exercise: not all units surveyed have an accounting system to track the specific expenses defined as composing R&D. The OECD regularly had to adjust or estimate national data to correct discrepancies. It also started a series called *Sources and Methods* documenting national differences with regard to OECD standards. It finally developed a whole system of footnotes, allowing for the construction of comparable data among member countries while black-boxing the data's limitations.

All in all, the GERD is not really a statistics on a national budget, but "a total constructed from the results of several surveys each with its own questionnaire and slightly [one could rather say significantly] different specifications". [29] Some data come from a survey (industry), others are estimated using different mathematical formulas (university), still other are proxies (government). For this reason: "The GERD, like any other social or economic statistic, can only be approximately true (…). Sector estimates probably vary from 5 to 15% in accuracy. The GERD serves as a general indicator of S&T and not as a detailed inventory of R&D (…). It is an estimate and as such can show trends (…)". [30]

[28] B. Godin (2008), *The Making of Statistical Standards: the OECD and the Frascati Manual, 1962-2002*, Project on the History and Sociology of STI Statistics, Montreal: INRS. Available at http://www.csiic.ca/PDF/Godin_39.pdf.
[29] D. L. Bosworth, R. A. Wilson and A. Young (1993), *Research and Development*, Reviews of United Kingdom Statistical Sources Series, vol. XXVI, London: Chapman and Hill, p. 29.
[30] Statistics Canada (2002), *Estimates of Total Expenditures on R&D in the Health Fields in Canada, 1988 to 2001*, 88F0006XIE2002007.

Nonetheless, according to a recent survey by the OECD Secretariat, GERD is currently the most cherished indicator among OECD member countries. [31] Over the last forty years, the indicator has been used for several purposes, from rhetorically displaying national performance to lobbying for more funds for science to setting policy targets. The OECD was responsible for this worldwide popularization of the indicator.

The OECD was also an ardent promoter of the **GERD/GDP** ratio as a policy target. It was Bernal who first suggested, in 1939, this type of measurement, which became the main indicator of science and technology: the research budget as a percentage of the national income. In the next decades, variants of the ratio took on names like research intensity, then technology intensity. The OECD made this statistic the ideal to which member countries would aim. In every OECD statistical publication, the indicator was calculated, discussed, and countries ranked according to it, because “it is memorable”, [32] and is “the most popular one at the science policy and political levels, where simplification can be a virtue”. [33]

The Frascati manual is entirely framed within an economic viewpoint. In the early 1960s, science was becoming recognized as a factor in economic growth. In order that science might optimally contribute to progress, however, science policies had to be developed. And to inform the latter, statistics were essential, or so thought the organization: “Informed policy decisions (…) must be based on accurate information about the extent and forms of investment in research, technological development, and scientific education”, argued the OECD’s Piganiol report. [34] “Provision for compilation of data is an indispensable prerequisite to formulating an effective national policy for science”.

What were the policy decisions for which data were so necessary? There were three, and all were framed within the vocabulary of neoclassical economics, even in evolutionary economists’ hands. The first was the **allocation of resources** to R&D, or what

[31] OECD (1998), *How to Improve the MSTI: First Suggestions From Users*, DSTI/EAS/STP/NESTI/RD (98) 9.
[32] OECD (1984), *Science and Technology Indicators*, Paris, p. 26.
[33] OECD (1992), *Science and Technology Policy: Review and Outlook 1991*, Paris, p. 111. The French translation reads as follows: “le plus prisé parmi les responsables de la politique scientifique et des hommes politiques, pour lesquels la simplification se pare parfois de certaines vertus” (p. 119).
[34] OECD (1963), *Science and the Policies of Government*, Paris: OECD, p. 24.

economists call the optimum level of resources: “Assessing what is in some sense the “right” or “optimum” level of allocation of resources”. [35] As discussed above, the GERD was developed to serve this end, and the ratio GERD/GDP became an indicator for policy targets.

The second policy decision was the **balance** between choices or priorities, or what economists call equilibrium. To many of those concerned, decisions about research funding were analyzed in terms of tensions between basic and applied research. To the OECD, statistics was the solution to the issue, and a system of classification for statistical breakdowns was proposed. The first edition of the Frascati manual suggested classifying R&D by dimensions. One of the central dimensions was concerned with economic sectors (industry, government, university, non-profit), as discussed above. Other classifications concerned each of the sectors.

Although each economic sector has its own classification, there is one more classification recommended in the manual, and it applies across all economic sectors. It concerns whether R&D is basic, applied or development, and this issue has been discussed for more than forty years at the OECD. The concept of basic research and its contrast with applied research has a long history that goes back to the nineteenth century, and the integration of the categories into taxonomies used for statistical surveys comes from the British scientists J. S. Huxley and J. D. Bernal. Since Condorcet, a magic number of 20 is often suggested as the percentage of R&D funds that should be devoted to basic research, and such a target was proposed by the OECD early on. [36]

We suggested that there were three policy decisions that required data, according to the OECD. The first was the allocation of resources to R&D. The second was balancing the budget. There was a third one, defined again according to neoclassical economics, namely determining the **efficiency**, or effectiveness of research. The first edition of the

[35] C. Freeman and A. Young (1965), *The Research and Development Effort in Western Europe, North America and the Soviet Union*, Paris: OECD, p. 15.
[36] Godin, B. (2003), Measuring Science: Is There Basic Research Without Statistics?, *Social Science Information*, 42 (1), pp. 57-90.

Frascati manual set the stage for measuring efficiency by using an input-output approach as a framework for science statistics: [37]

Input → Research activities → Output

Certainly the manual was entirely concerned with proposing standards for the measurement of inputs. But this was only a first stage. [38] Despite this focus, the manual discussed output and inserted a chapter (section) specifically dedicated to its measurement because "in order really to assess R&D efficiency, some measures of output should be found". [39] From its very first edition, then, the Frascati manual suggested that a complete set of statistics and indicators, covering both input and output, was necessary in order to properly measure science. Since then, the OECD has developed a whole family of manuals covering both input and output:

The OECD R&D Family of Manuals

(First edition)

1961	*The Measurement of Scientific and Technical Activities: Proposed Standard Practice for Surveys of Research and Development* (Frascati manual).
1990	*Proposed Standard Practice for the Collection and Interpretation of Data on the Technological Balance of Payments*.
1992	*Proposed Guidelines for Collecting and Interpreting Technological Innovation Data* (Oslo manual).
1994	*Data on Patents and Their Utilization as Science and Technology Indicators*.
1995	*Manual on the Measurement of Human Resources in Science and Technology* (Canberra manual).

[37] Godin, B. (2007), Science, Accounting and Statistics: the Input-Output Framework, *Research Policy*, 36 (9), pp. 1388-1403.
[38] OECD (1962), The Measurement of Scientific and Technical Activities: Proposed Standard Practice for Surveys of Research and Development, *op. cit.* p. 11.
[39] *Ibid*.

Defining Science with Statistics

Four elements have characterized the official definition of science over the twentieth century. [40] First, science has been defined and measured by officials based on the concept of "research". This is a purely social construction, since science could also be defined otherwise than as activity, or research. Scientists and philosophers have long defined science by its output (knowledge) and method, economists have defined it as information, and sociologists have defined it by its institutions and practices. Early officials' definitions also varied. Until recently, the USSR and the communist countries, for example, used a broader definition, in which science covered more than research, i.e.: covered areas excluded from the OECD definition of research since they were qualified as related scientific activities, for example scientific information and standardization. UNESCO, for its part, developed the concept of scientific and technological activities, which included research, education and related scientific activities.

Defining science as research is due to the institutionalization of research as a major phenomenon of the 20th Century. By the 1960s, most large organizations have recognized research as a contributor to economic growth, performance, and innovation, and many organizations were devoting an increasing share of their budget to these activities. Hence the need for a better understanding of what was happening and for measuring the efforts (as a first step in the measurement of science).

However, this definition owes to a second factor, namely accounting and its methodology. There are activities that are easily measurable and others that are not. There are activities for which numbers are available, and others for which they are not. There are activities that can be identified and distinguished easily, and some that in practice are difficult to separate. Officials chose to concentrate on the more easily measurable (R&D), for methodological reasons having to do with accounting (costs) and its measurement: research activities rather than research outputs (or knowledge), research activities rather than (research plus) related scientific activities, research and

[40] B. Godin (2008), *What is Science? Defining Science by the Numbers, 1920-2000* , Project on the History and Sociology of STI Statistics, Montreal: INRS. Available at http://www.csiic.ca/PDF/Godin_35.pdf.

development rather than research exclusively, and purely systematic research rather than (systematic and) *ad hoc*. Let's look at these choices.

The second characteristic of science as defined by governments and its statistics is R&D. Research is defined essentially as R&D, where "D", for development, corresponds to over two-thirds of expenses. Development is composed of several activities like scale activities, pilot plants and design. It is an important category of taxonomies on research. Since the survey on industrial research by accountant R. N. Anthony from Harvard University, conducted for the US Department of Defense in the early 1950s, research is defined as composed of three categories: [41] basic research, applied research and development. Development got into R&D for many reasons, among them because of its importance in industrial (and military) research and because of the difficulty of separating (and budgeting) development from other activities like research proper. It also owes its presence in the definition to the priority that technological development had on the science policy agenda. [42]

However, in the 1960s, in light of increasing expenditures on R&D as reported in official statistics, particularly military R&D, some began questioning what really goes into statistics on research. David Novick, from RAND Corporation, suggested: "we should stop talking about research and development as though they were an entity and examine research on its own and development as a separate and distinct activity." [43] The rationale for this suggestion was one provided by S. Kuznets and J. Schmookler a few years earlier: "development is a job of adjustment (…); it is not original invention"; [44] "while the problems dealt with in development are non-routine, their solution often does not

[41] D. C. Dearborn, R. W. Kneznek and R. N. Anthony (1953), *Spending for Industrial Research, 1951-1952*, Division of Research, Graduate School of Business Administration, Harvard University.
[42] B. Godin (2006), Research and Development: How the "D" Got Into R&D, *Science and Public Policy*, 33 (1), pp. 59-76.
[43] D. Novick (1965), The ABC of R&D, *Challenge*, June, p. 13. See also: D. Novick (1960), What do we Mean by R&D?, *Air Force Magazine*, October, 114-118.
[44] S. Kuznets (1962), Inventive Activity: Problems of Definition and Measurement, in NBER, *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton: Princeton University Press, p. 35.

demand the creative faculty which the term invention implies". [45] All three authors lost this argument.

The third characteristic of the official definition of research is the idea of "systematicness". Industrial research underwent expansion after World War I. Most big firms became convinced of the necessity to invest in research and began building laboratories for the purpose of conducting research: research had to be "organized and systematized". The issue of "systematically" organizing industrial research was on every manager's lips. This is the rationale behind the official definition of research. Research is organized research, i.e.: laboratory research. The meaning spread rapidly through surveys of research activities.

It was the NSF and the OECD that generalized this concept of research. Two aspects of this conception deserve analysis. First, the meaning of systematic used in defining research – and the statistics based thereon – has drifted from an emphasis on the *scientific method* to an emphasis on *institutionalized* research. This drift was closely related to the (modern) instrument used for measuring research, namely the survey, and to that instrument's limitations. Second, the definition had important consequences on the numbers generated, the most important one being the undercounting of research. Let us discuss both aspects.

The origins of this state of affairs are the industrial survey and its influence on the whole methodology of questionnaires, including questionnaires for surveying government and university research. The main link here was US accountant R.N. Anthony. In the survey he conducted for the Department of Defense, Anthony showed that firm size was one of the main variables explaining R&D investment. Consequently, he suggested: [46]

> The fact that there are almost 3,000 industrial research organizations can be misleading. Most of them are small. (…) Over half employ less than 15 persons each, counting both technical and non-technical personnel. Many of these small laboratories

[45] J. Schmookler (1962), Comment on S. Kuznets' paper, in NBER, The Rate and Direction of Inventive Activity: Economic and Social Factors, *op. cit*. p. 45.

[46] R. N. Anthony and J. S. Day (1952), *Management Controls in Industrial Research Organizations*, Boston: Harvard University, pp. 6-7.

> are engaged primarily in activities, such as quality control, which are not research or development.
>
> [Therefore] this report is primarily concerned with industrial laboratories employing somewhat more than 15 persons.

Hence, research was thereafter equated with systematized research or large organizations with dedicated laboratories. This rationale soon came to be related to another one: the costs of conducting a survey. Because there are tens of thousands of firms in a country, units surveyed have to be limited to manageable proportions. This was done by introducing a bias in industrial surveys: the survey identified all major R&D performers, that is big firms with laboratories (or "organized" research) and surveyed them all, but selected only a sample of smaller performers, when they selected any. This decision was also supported by the fact that only big firms had precise book-keeping practices on R&D, since the activity could be located in a distinct and formal entity, the laboratory. Thus, an important impact of the official concept of research was the undercounting of R&D and, therefore, a failure to support some performers in science policies. Authors have measured four times as many man/years devoted to R&D in small and medium sized companies than what had been reported in government surveys. The reason offered for the differences was that small and medium sized companies tend to conduct R&D in an informal way ("unorganized", some would say), rather than on a continuous basis or in a department of the firm exclusively devoted to R&D.

The fourth and last aspect of the official concept of research is the exclusion of a certain type of activities, namely those called related scientific activities. The choice made was to separate research from other (routine) activities, however indispensable they may be to research: planning and administration, expansion of R&D plant, data collection, dissemination of scientific information, training, and testing and standardization. In fact, firms had accounting practices that did not allow these activities to be easily separated.

The decision to concentrate on research, or R&D, was not without its opponents. We owe to UNESCO the development of a more inclusive definition of science. First, with regard to related scientific activities, the fact that the organization was devoted to educational and cultural development as much as economic development explains its interest in related scientific activities. Also, the fact that the organization was dominated by scientists, not economists as was the case at OECD, was also an influential factor in defining science differently. According to that organization, surveying national science and technology "should not be limited to R&D but should cover related scientific and technological activities (…). Such activities play an *essential* part in the scientific and technological development of a nation". [47]

UNESCO's interest in related scientific activities was the consequence of its basic goal of extending standardization beyond industrialized (i.e.: OECD) countries. What was peculiar to eastern countries at the time was the fact that R&D was not designated as such. The USSR, for example, put all its statistics on science and technology under the heading "science". In attempting to accommodate eastern Europe, however, UNESCO's efforts were guided as much by the desire to generate a larger range of standardization than the OECD as by an interest in related scientific activities *per se*. But the program for including eastern Europe failed, and UNESCO never collected data on related scientific activities. Why? The reasons are many.

First, UNESCO itself concentrated on R&D. The activity was said to be easier to locate and to measure, and had the virtue of being an "exceptional" contribution to science and technology. R&D was perceived as a higher order of activity. The second reason that UNESCO never pursued work on related scientific activities was linked to the fact that, in the end, few countries were interested in these activities. But the main reason that UNESCO failed in its efforts to measure related scientific activities was that the United States left the organization in 1984, accusing UNESCO of ideological biases. The decision had a considerable impact on the UNESCO Division of Statistics in terms of financial and human resources.

The concept of "scientific and technological activities" was the second effort of UNESCO to broaden the definition and measurement of science, and would become the basis of UNESCO's philosophy of science measurement:

[47] UNESCO (1970), *Manual for Surveying National Scientific and Technological Potential*, NS/SPS/15, Paris: UNESCO, p. 21.

> Broadening of the scope of science statistics is particularly appropriate to the conditions of most of the developing countries which are normally engaged in more general scientific and technological activities, rather than R&D solely. [48] In developing countries proportionally more resources are devoted to scientific activities related to the transfer of technology and the utilization of known techniques than to R&D per se. [49]

According to the UNESCO recommendation, adopted by member countries in 1978, scientific and technological activities were composed of three broad types of activities: R&D, scientific and technical education and training, and scientific and technological services (or related scientific activities) (Figure 1). [50] The UNESCO recommendation was short-lived. In 1986, the director of the UNESCO division of statistics on science and technology concluded that "Due to considerable costs and organizational difficulties, the establishment of a system of data collection covering at once the full scope of scientific and technological services and S&T education and training in a country has been considered not practicable".

Conclusion

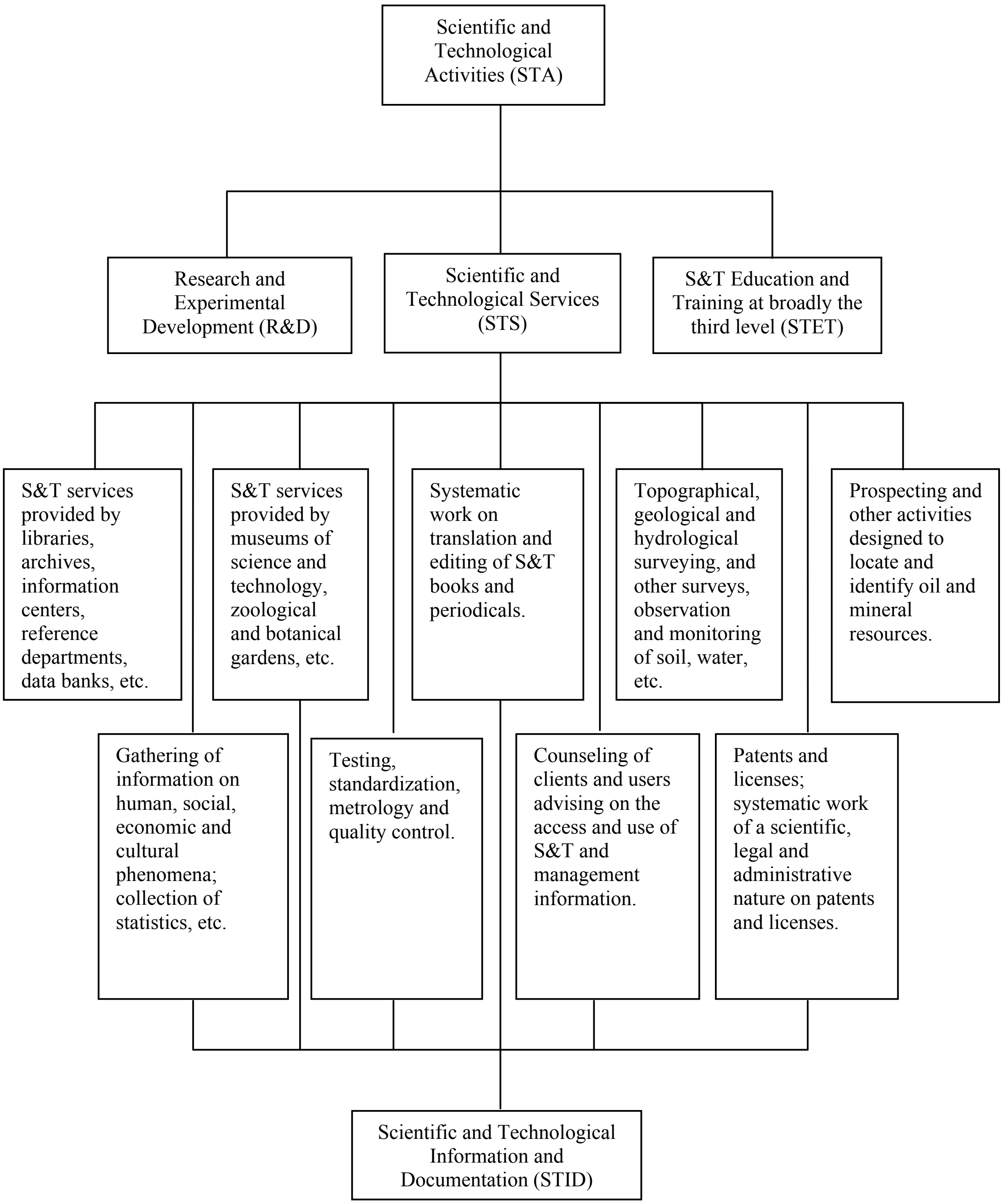
The measurement of science is a fascinating episode in the history of science: it is witness of ideological, political, social and economic interests. From the start, measuring the number of scientists rather than other aspects on science had to do with the context of the time. To many people, the stock of the population and the quality of the race was deteriorating, and those groups that contributed more to civilization, namely eminent men, including scientists, were not reproducing enough and had insufficient incentives and recognition. The "unfits" were far more productive – and some suggested policies for sterilizing them. This gave rise to the idea of measuring the number of available scientists, the size of the scientific community and the social conditions of scientists as researchers.

---

[48] UNESCO (1969), *Science Statistics in Relation to General Economic Statistics: Current Status and Future Directions*, UNESCO/COM/CONF.22/2, p. 9.
[49] UNESCO (1972), *Considerations on the International Standardization of Science Statistics*, COM-72/CONF.15/4, p. 14.
[50] UNESCO (1978), *Recommendation Concerning the International Standardization of Statistics on Science and Technology*, Paris: UNESCO.

Figure 1.

S&T Activities (UNESCO)



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. 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), became the most cherished indicator.

The main consequence of such an orientation for statistics was twofold. First, statistics came to be packaged in an accounting framework. Statistics on science concentrated on costs, aligning themselves with the System of National Accounts, and were collected within an input/output approach. Most current indicators are economic in type: expenditures on research, output such as technological balance of payments, patents, high-technology products, marketed innovation, etc. The second consequence was a focus on economic growth and productivity. Certainly, the concept of scientific productivity in science arose from scientists themselves. In Galton's hands, productivity meant reproduction: the number of children a scientist had, or the number of scientists a nation produces. Then, in the twentieth century, scientific productivity came to mean the quantity of output of a scientific or technological type (papers, patents), and later economic (labour or multifactor) productivity, or impacts of science on economic growth. [51]

Today, it is the organizations (and the economic sector to which they belong) that are measured, above all firms (think of the innovation surveys), and not the people from society who are supposed to benefit from science. In spite of decades, even centuries, of discourses on the social benefits of science, you would look in vain for systematic indicators on the social side of science. In fact, to "accounting", the economics is what is significant, what is rendered visible and what becomes imperative for action. The social

[51] B. Godin (2009), The Value of Science: Changing Conceptions of Scientific Productivity, 1869-circa 1970, *Social Science Information*, Forthcoming.

is the residual and is relegated to the periphery. [52] The culture of numbers is in fact the cult of (economic) efficiency.

[52] A. G. Hopwood (1984), Accounting and the Pursuit of Efficiency, in A. G. Hopwood and C. Tomkins (eds.), *Issues in Public Sector Accounting*, Oxford: Philip Allan, pp. 167-187.