Metadata:
How Footnotes Make for Doubtful Numbers

Benoît Godin

3465, rue Durocher
Montréal (QUE)
Canada H2X 2C6

Project on the History and Sociology of S&T Statistics

Working Paper No. 10

2001

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*.

Metadata:
How Footnotes Make for Doubtful Numbers

Objectivity is a central characteristic of science. Whether we consider science from the perspective of scientists or philosophers, objectivity defines its very nature. It took hundreds of years to develop the means by which the evidence that defines objectivity came to be accepted. [1] It began in the 17th Century with the testimony of witnesses on experiments performed in "public" spaces, and coalesced into the "virtual witnessing" of the scientific paper with its detailed instructions for replicating experiments. [2] Today, facts and data are acceptable forms of evidence in the natural sciences. Similarly, the social and human sciences are also founded on "facts": the survey is a major source of data in the social sciences, while the archives and documents are the principal sources of data in the humanities. [3]

In all these scientific ventures, whether natural or social, footnotes served two particular functions. [4] Firstly, footnotes indicated sources of evidence: they attest to facts. They offer empirical support for stories told and arguments presented. Secondly, they seek to persuade the reader that the scientist has done an acceptable amount of work. They confer authority on a writer by establishing his or her seriousness, expertise and credibility. [5]

Besides being a source of evidence and a tool of persuasion however, I submit that footnotes sometimes serve a third function: they disclose the data's weaknesses in the process of qualifying its strengths. Footnotes "make clear the limitations of their own theses even as they try to back them up. (…) Footnotes prove that [the narrative] is a historically contingent product, dependent on the forms of research, opportunities, and

[1] P. Dear (1995), *Discipline and Experience: The Mathematical Way in the Scientific Revolution*, Chicago: University of Chicago.
[2] S. Shapin and S. Schaeffer (1985), *Leviathan and the Air Pump: Hobbes, Boyle and the Experimental Life*, Princeton: Princeton University Press.
[3] B.S. Shapiro (2000), *A Culture of Facts, England 1550-1720*, Ithaca: Cornell University Press; M. Poovey (1998), *A History of the Modern Fact*, Chicago: University of Chicago Press.
[4] A. Grafton (1997), *The Footnote: A Curious History*, Cambridge (Mass.): Harvard University Press.
[5] N.G. Gilbert (1977), Referencing as Persuasion, *Social Studies of Science*, 7: 113-122.

states of particular questions". [6] As we shall see later, footnotes often betray an antirealist rhetoric: authors acknowledge the data's shortcomings in footnotes while ignoring them in the body of the text.

This paper examines the treatment of limitations in science and technology statistics. In 1963, Member countries of the OECD unanimously approved a methodological manual for standardizing science and technology statistics. [7] The aim of the manual was to harmonize national practices in conducting R&D surveys. Several editions later, however, there are still many methodological differences that continue to hamper international comparisons. As L.A. Seymour noted nearly forty years ago: "as one proceeds from comparisons of two countries to several countries, each with an independent system of data collection, the complications multiply so rapidly that any thought of further extension seems presently unfeasible". [8]

As a result of international non-comparability, the OECD regularly corrected the national data or simply added footnotes to inform the user of discrepancies in the series. It is a well-known fact that measurement abstracts only few properties to better compare objects. Eliminating details permit the construction of tables that would not otherwise exist: statistical tables present standardized data that exclude national specificities - by relegating them to footnotes. In the present paper, I want to show how, for those who bother to read the footnotes, the data often appear as mere tendencies only, if not veritable fictions.

The first part of the paper presents the problems and difficulties that national statisticians faced in analyzing R&D data before the OECD's involvement in science and technology statistics. There were three kinds of problems: the definition of concepts, the demarcation between scientific and non-scientific activities, and the sample of the population surveyed. The second part analyzes the problems of comparability that led the OECD to standardize

[6] A. Grafton*, op. cit.,* p. 23.
[7] OECD (1963), *The Measurement of Scientific and Technical Activities: Proposed Standard Practice for Surveys of Research and Development*, Paris.
[8] L.A. Seymour (1963), *Problems of International Comparisons of R&D Statistics*, OECD, DAS/PD/63.3, p. 4.

practices between countries. Despite these efforts, differences persisted and gave rise to the development of tools, like the footnote, for minimizing national specificities.

Living with Differences

In 1959, J. Perlman, head of the Office of Special Studies at the NSF, praised his staff for having introduced two innovations in the field of R&D surveys: the regularity and comprehensiveness of surveys, through the periodic coverage of all economic sectors (government, industry, university), and the analysis of intersectorial flows of funds. [9] But he failed to mention a third innovation: the NSF had put an end to variations in the methodology of R&D surveys, at least in the United States.

Prior to the 1960s, there were three problems that prevented statisticians from comparing surveys, drawing historical series or even believing in the numbers. The first problem concerned definitions of research. Two situations prevailed. Firstly, more often than not, there was no definition of research at all, as was the case in the US National Research Council (NRC) repertory of industrial R&D. The first edition reported using a "liberal interpretation" that let each firm decide which activities counted as research: "all laboratories have been included which have supplied information and which by a liberal interpretation do any research work". [10] Consequently, any studies that used NRC numbers, like those by Holland and Spraragen [11] and the US Work Progress Administration (WPA) [12] were of questionable quality: "the use of this information [NRC data] for statistical analysis has therefore presented several difficult problems and has necessarily placed some limitations on the accuracy of the tabulated material". [13]

[9] NSF (1959), *Methodological Aspects of Statistics on Research and Development: Costs and Manpower*, NSF 59-36, Washington, pp. 2-3.
[10] National Research Council, *Research Laboratories in Industrial Establishments of the United States of America*, Bulletin of the NRC, vol. 1, part 2, March 1920, p. 45.
[11] M. Holland and W. Spraragen (1933), *Research in Hard Times*, Division of Engineering and Industrial Research, National Research Council, Washington.
[12] G. Perazich and P.M. Field (1940), *Industrial Research and Changing Technology*, Work Projects Administration, National Research Project, report no. M-4, Pennsylvania: Philadelphia.
[13] G. Perazich and P.M. Field (1940), *op. cit*., p. 52.

The US National Resources Planning Board used a similar practice in its survey of industrial R&D in 1941: the task of defining the scope of activities to be included under research was left to the respondent. [14] In Canada as well, the first study by the Dominion Bureau of Statistics contained no definition of research. [15]

The second problematic situation regarding definitions was the use of categories of research in *lieu* of a precise definition. Both the Bush [16] and the Steelman reports, [17] as well as the British DSIR survey [18] suggested categories that resembled each other (basic, applied, development) – but that were never in fact the same. [19] As a rule, these categories served to help respondents decide what to include in the questionnaire, but they were rarely used to calculate statistical breakdowns. In fact, interest in statistics on basic research began only after the NSF started paying it serious attention in the 1950s. [20] Finally, others simply refused to use such categories, such as the National Resources Committee, because of the intrinsic connections between basic and applied research that seemed to prevent any clear distinctions from being made. [21]

The situation improved thanks only to the NSF and the OECD. Some precise definitions were indeed proposed as early as 1938 by the National Resources Committee in its survey of government R&D, [22] but they were limited to government R&D and aroused controversy for including the social sciences. The Canadian Government also suggested (influential) definitions, like that of "scientific activities", as early as 1947, [23] but a standard definition would have to wait until the creation of the NSF and the OECD. The idea of

[14] National Resources Planning Board (1941), *Research: A National Resource (II): Industrial Research*, Washington: USGPO, p. 173.
[15] Dominion Bureau of Statistics (1941), *Survey of Scientific and Industrial Laboratories in Canada*, Ottawa.
[16] V. Bush (1945), *Science: The Endless Frontier*, North Stratford: Ayer Co. Publishers, 1995, pp. 81-83.
[17] J.R. Steelman (1947), *Science and Public Policy*, New York: Arno Press, 1980, pp. 300-301.
[18] DSIR (1958), *Estimates of Resources Devoted to Scientific and Engineering R&D in British Manufacturing Industry, 1955*, London, p. 8.
[19] See: B. Godin (2000), *Measuring Science: Is There Basic Research without Statistics?*, Montreal :OST.
[20] Ibid.
[21] National Resources Committee (1938), *Research: A National Resource (I): Relation of the Federal Government to Research*, Washington: USGPO, p. 6.
[22] National Resources Committee (1938), *op.cit.*, p. 62.
[23] Department of Reconstruction and Supply (1947), *Research and Scientific Activity: Canadian Federal Expenditures 1938-1946*, Government of Canada: Ottawa, pp. 11-13.

"systematicity" would thereafter define research: [24] "*systematic*, intensive study directed toward fuller knowledge of the subject studied and the *systematic* use of that knowledge for the production of useful materials, systems, methods, or processes". [25]

The second problem of pre-1960s R&D surveys, closely related to the problem of definition, concerned the demarcations of research and non-research activities. The main purpose of both the Harvard Business School study and the US Bureau of Labor Statistics survey, two influential studies of the early 1950s, was to propose a definition of R&D and to measure it. Two problems were identified: there were too many variations on what constituted R&D, so they claimed, and too many differences among firms on which expenses to include in R&D. [26] Although routine work was almost always excluded, there were wide discrepancies at the frontier between development and production, and between scientific and non-scientific activities: testing, pilot plant, design, and market studies were sometimes included in research and at other times not (Annex 1). Indeed, companies had accounting practices that did not allow these activities to be easily separated. K. Arnow, of the NSF, summarized the problem as follows:

> Even if all the organizations responding to the NSF's statistical inquiries shared, by some miracle, a common core of concepts and definitions, they might still not be able to furnish comparable data, since they draw on a diversity of budget documents, project reports, production records, and the like for estimating R&D expenditures. [27]

According to R.N. Anthony, author of the Harvard Business School survey, accounting practices could result in variations of up to 20% for numbers on industrial R&D. [28] Both the US Bureau of Labor Statistics [29] and the NSF also believed that only better accounting practices could correct such errors. Nevertheless, the Harvard Business School [30] and the

[24] B. Godin (2000), *Defining R&D: Is Research Always Systematic?*, Montreal: OST.
[25] National Science Foundation (1953), *Federal Funds for Science*, Washington, p. 3.
[26] 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, p. 91.
[27] K. Arnow (1959), National Accounts on R&D: The NSF Experience, in NSF, *Methodological Aspects of Statistics on Research and Development: Costs and Manpower*, NSF 59-36, Washington: 58.
[28] R.N. Anthony (1951), *Selected Operating Data: Industrial Research Laboratories*, Harvard Business School, Division of Research, Boston, p. 3.
[29] H. Wood, Some Landmarks in Future Goals of Statistics on R&D, in NSF (1959), *op. cit*: 52; NSF (1960), *op. cit*, p. 99.
[30] D.C. Dearborn, R.W. Kneznek and R.N. Anthony (1953), *op.cit*, pp. 43-44, 92.

NSF [31] both insisted to develop a whole series of specifications for defining and delimiting measurable activities. The first NSF industrial R&D survey included pilot plant, design, laboratory scale and prototypes in its definition of research; and it excluded market and economic research, legal work, and technical services (minor adaptations, licenses, advertising, patents, and exploration). In the following decades, the OECD improved and standardized these demarcations through the Frascati manual, [32] and it concentrated on measuring research activities as such. But related scientific activities have rarely been systematically measured by any organization. [33]

A third and final problem of early R&D surveys concerned the sample of the population under study. We have seen how the NRC repertory was open to all firms who agreed to complete the questionnaire: "the NRC surveys were designed for the purpose of compiling a series of directories of research laboratories in the United States. The schedules were therefore sent out without instructions which would have been necessary had it been intended to use the data for purposes of statistical analysis". [34] When the statisticians finally begun addressing the problem, however, their methodologies differed: some limited the survey to distinct laboratories, [35] others sent the questionnaire on a consolidated company basis, [36] still others concentrated on big firms to "speed up results". [37] There were no real standards.

All in all, the absence of norms made survey comparisons impossible before the sixties, which resulted in statistics that were often of limited value. Steelman wrote that it was "not possible to arrive at precisely accurate research expenditures" because of three limitations: 1) variations in definition, 2) accounting practices, and 3) the absence of a clear division

[31] National Science Foundation (1953), *Federal Funds for Science: Federal Funds for Scientific R&D at Nonprofit Institutions 1950-1951 and 1951-1952*, Washington, p. 16.
[32] B. Godin (2001), *Neglected Scientific Activities: The (Non) Measurement of Related Scientific Activities*, Montreal: OST.
[33] The exception was: D.C. Dearborn, R.W. Kneznek and R.N. Anthony (1953), *op. cit*.
[34] G. Perazich and P.M. Field (1940), *op. cit*., p. 52.
[35] R.N. Anthony (1951), *op. cit*, p. 42.
[36] R.N. Anthony (1953), *op. cit*., p. 43.
[37] Dominion Bureau of Statistics (1956), *Industrial Research-Development Expenditures in Canada, 1955*, Ottawa, p. 22.

between science and other research activities. [38] Similarly, the NSF admitted that the industrial R&D surveys it conducted before 1957 were not comparable to those it conducted after. [39] However, statisticians commonly defended the data despite its limitations and biases: [40]

> These difficulties have resulted in wide variations in the analyses of research expenditures which have been published during the last decade, but they do not affect orders of magnitude or the general trends on which policy decisions rest. [41]
>
> It is sometimes usually difficult for budget officers to draw the line between research and administration (…). Nevertheless (…) the results are believed significant, at least for compilation into totals for Government as a whole, if not for detailed inter-agency comparisons. [42]
>
> The data collected in this way are, of course, not complete. Many organizations doing research have not been reached, nor are the returns received always comparable. However, it is believed that the coverage is quite adequate to yield a representative and qualitatively correct picture of present day industrial research. [43]

Similarly, the Canadian Dominion Bureau of Statistics once asserted: "although the records of some respondents did not follow the definitions (…) it is felt that any variations in interpretation of the type of data to be included in the questionnaire were not significant enough to make any appreciable difference in the published data". [44] The US Bureau of Labor Statistics also stated: "despite these limitations, the findings of this survey are believed to give a satisfactory general picture. (…) The reader should, however, bear in mind the approximate nature of the figures". [45]

The early NSF surveys presented the same rhetoric: [46] "While the amounts reported do not correspond exactly to what was actually received by nonprofit institutions in either of the

[38] J.R. Steelman (1947), *op. cit*, pp. 73, 301.
[39] NSF (1960), *Funds for R&D: Industry 1957*, NSF 60-49, Washington, pp. 97-100
[40] For similar rhetorical strategies on the part of academic researchers, see for example: E. Mansfield (1991), Academic Research and Industrial Innovation, *Research Policy*, 20: 11; E. Mansfield (1972), Contribution of R&D to Economic Growth in the United States, *Science*, 175, 4 February, pp. 480 and 482.
[41] [41] J.R. Steelman (1947), *op. cit*, pp. 73.
[42] National Resources Committee (1938), *op. cit*., p. 63.
[43] National Resources Planning Board (1941), *op. cit*., p. 173.
[44] Dominion Bureau of Statistics (1956), *op. cit*., p. 22.
[45] US Department of Labor, Bureau of Labor Statistics, Department of Defense (1953), *Scientific R&D in American Industry: A Study of Manpower and Costs*, Bulletin no. 1148, Washington, p. 46.
[46] National Science Foundation (1953), *op. cit*, p. 5.

two years covered, they do represent orders of magnitude from which generalizations may be drawn". [47] This was in fact the view, as we shall see below, that would be adopted by the OECD R&D statistical unit: because they were imperfect, R&D statistics should be interpreted as tendencies or trends only, as suggested by the titles of most of its analytical publications on R&D data of the seventies and in the following citation from a study on indicators of high-technology: [48]

> The findings must be viewed as indicating principally broad trends. The data only gives a partial representation of the role played by high-technology processes and products in international trade. This is due to the intrinsic limitations of the method followed in preparing the figures.

All these positions had nothing to do with statistics *per se*, as the second OECD ad hoc users group on R&D statistics would soon report: [49] "Most doubts about national data have nothing to do with standard deviations and sampling theory but concern whether the data in a response reflect the reality in the firm, or institute or university laboratory involved. There are a very large number of reasons why they may not, including the ability to distinguish R&D conceptually from other activities, unsuitable accounting systems, loosely drafted national questionnaire and fiscal or juridical incentives to answer inaccurately". [50]

While most studies minimized the data's limitations, some were also cautious – usually in technical notes - about the statistics. The NSF, for instance, regularly calculated the margin of error of its surveys, estimating in 1960 that the difficulties associated with the concept of research were responsible for discrepancies of up to 8%. [51] It was also very careful in interpreting research categories (or character of work): as early as 1956, the NSF stated that its estimates were maximums at best because of differing interpretations and accounting

[47] This was a problem that would occupy State statisticians for years: the difference between the amount of performed versus financed R&D.
[48] OECD (1984), *An Initial Contribution to the Statistical Analysis of Trade Patterns in High Technology Products*, DSTI/SPR/84.66, p. 1.
[49] OECD (1978), *Report of the Second Ad Hoc Review Group of R&D Statistics*, SPT (78) 6: 17-18.
[50] As an example, the actual rate of non-responses in surveys has to do with the difficulty of the concepts and questions. Non-responses are not randomly distributed but biased with respect to certain characteristics of the population and the questionnaire. See: M. Akerblom (2001), *Develop Proposed Standard Methodology for R&D Surveys*, DSTI/EAS/STP/NESTI (2001) 14/PART2, p. 10.
[51] NSF (1960), op.cit., p. 97.

practices. [52] In 1960, after its third survey of industrial R&D, the NSF noted that "the accounting systems of some companies are not set up to yield accurate data on R&D. (…) Companies find it difficult to define boundaries differentiating basic from applied research, applied research from development, or development from production, testing or customer-service work". [53] The NSF would consequently, modify and soften the definition of basic research for industry [54] and develop methods for estimating corporate basic research because of non-responses, which reached 40% in the 1980s. [55] But one thing was sure: it would never abandon the concept of basic research itself. Former NSF director D.N. Langenberg explained: the NSF "must retain some ability to characterize, even to quantify, the state of the balance between basic and applied research across the Foundation. It must do so in order to manage the balance properly and to assure the Congress and the scientific and engineering community that it is doing so". [56]

Comparing the Incomparable

The NSF put an end to variations in the methodology of R&D surveys because it monopolized and imposed its own standards in the field of science and technology statistics in the United States. The OECD tried replicating the experience on an international scale in the sixties, and achieved a certain measure of success.

In 1965, C. Freeman and A. Young produced a study that measured R&D gaps between OECD member countries. [57] It was the first attempt to compare R&D internationally, [58] and represented the state of the art in R&D statistics in 1962 (the latest year for which there

[52] NSF (1956), *Science and Engineering in American Industry: Final Report on a 1953-54 Survey*, NSF 56-16, Washington, pp. 18, 48.
[53] NSF (1960), *op. cit,.* pp. XII-XIII.
[54] See B. Godin, *Measuring Science*, op. cit.
[55] NSF (1990), *Estimating Basic and Applied R&D in Industry: A Preliminary Review of Survey Procedures*, NSF 90-322, Washington.
[56] D.N. Langenberg (1980), *Memorandum for Members of the National Science Board*, NSB-80-358, Wasington, p. 4.
[57] C. Freeman and A. Young (1965), *The Research and Development Effort in Western Europe, North America and the Soviet Union: An Experimental International Comparison of Research Expenditures and Manpower in 1962*, Paris: OECD.

were data) - one year before the first international survey based on the Frascati manual. The study included an appendix called *Sources and Methods* that dealt with national differences in statistics and limitations of data, and that would serve as the model for the OECD "Sources and Methods" series. The study revealed "big differences in definition and coverage" among countries. [59] These concerned, to name but a few, the treatment of: the social sciences, capital expenditures, funds to R&D abroad, related scientific activities, government funding of industrial R&D, and scientific personnel (see Annex 2 for details).

These were precisely the differences that the Frascati manual, adopted in 1963 by OECD Member countries, was designed to eliminate. [60] The manual's proposed standards were mainly concerned with four topics. Firstly, norms were proposed for defining research as "systematic" search [61] and the three major categories of research (basic/applied/development). [62] Secondly, activities were demarcated for statistical inclusion or exclusion: [63] research/related scientific activities, development/production, research/teaching. Thirdly, sectors (university, government, industry, non-profit) were precisely delineated. Finally, standards were suggested for surveying the units of research.

As a result of the manual, the first international survey of R&D was conducted in seventeen countries in 1963-64. The results were published in 1967 in a small booklet that discussed very briefly, among other things, limitations before any number was presented, but at the same time repeatedly mentioned that this did not affect the results. The following year, a huge document was published, entirely "designed to clarify national particularities and certain problems of a general nature which impede international comparability". [64] The fact that the methodological notes were published separately indicated that they were rarely read. The document, nevertheless, included three sets of limitations and notes. Firstly, the

[58] See also: C. Freeman (1962), Research and Development: A Comparison Between British and American Industry, *National Institute Economics Review*, 20, May: 21-39.
[59] C. Freeman and A. Young (1965), *op.cit*. p. 19.
[60] OECD (1963), *op.cit*.
[61] B. Godin (2001), Defining Research, *op. cit*.
[62] B. Godin (2000), Measuring Science, *op. cit*.
[63] B. Godin (2000), Neglected Scientific Activities, *op. cit*.
[64] OECD (1968), *A Study of Resources Devoted to R&D Member Countries in 1963/64: Statistical Tables and Notes*, Paris, p. 3.

introduction presented an overall evaluation of the data (pp. 17-26), and was followed by a discussion of the problem of exchange rates (27-32). Secondly, a series of notes accompanied each table: footnotes (ex.: p. 37) and endnotes (ex. : p. 56), which latter would become what is now known as the "standard footnotes". Finally**,** the statistics for each sector were preceded by general notes (ex.: p. 73), and by notes for each of the surveyed countries (ex. : p. 77). The same model was subsequently used in each biennial publication of the survey (ISY).

The technical document identified two general sources of errors in making international comparisons:

Reliability of data

1. Omission of important R&D performers
2. Non-responses of important R&D performer,
3. Inaccurate grossing up, extrapolations, estimations.

Variations in concepts and definitions

4. Standards (Frascati manual) not followed
5. Standards interpreted differently
6. No standards in Frascati manual

This was just the beginning of a long series of incompatible national practices that took on more and more importance: breaks in series caused by changes in definitions of concepts; differing periods of the national surveys (civil, academic, budgetary) and variations in their frequency (1, 2, 3 years; alternating sectors); non-comparable samples, unreliable estimates (of non-responding or non-surveyed units); and differing coverage of sectors - all of which led the OECD to make estimations of national data with increasing frequency. [65]

> In cases where differences from the prescribed standard occurred, the Secretariat worked closely with national authorities to rectify discrepancies. Inevitably, however, certain

[65] *Ibid*, p. 15.

> problems of comparability persisted. The tables for each sector are therefore preceded by an introductory note and by detailed country-by-country notes covering all know differences between national definitions and survey methods.

The practice of including notes took on particular importance in the 1980s, because of two events. Firstly, the second OECD users group on R&D statistics dealt with the unacceptably large diffusion lags in the publication of data, [66] and also with the methodological notes: [67]

> The single most prevalent concern expressed by users about the adequacy of OECD R&D data resolved around the extent to which data from different countries are comparable, both with respect to accuracy and definition of scope.
>
> On one level, it can be said that current publications provide an answer to this concern inasmuch as all data series are accompanied by often extensive country notes. However, it is evident that many users do not give detailed attention to such notes and so a better method of dealing with this problem is needed.

At the end of the 1990s, there is the same issue of "protecting users from themselves since many users, especially those close to policy makers, do not understand, care about or choose to read about the statistical problems which are described in footnotes or in the sources and methods of official database publications". [68] The group therefore suggested that: "the Secretariat attempt to summarize the country notes with a view to highlighting those series of data for which problems of comparability are most acute". [69] The recommendations led to the development of "standard footnotes" for each statistical publication (Annexes 3 and 4).

The second event that increased the development of notes was a series of in-depth R&D studies by the OECD Science and Technology Indicators Unit (STIU) in the 1970s (Annex 5). Three groups of problems were revealed by the analysis of sectorial R&D (university, industry, and government).

[66] On the sources of the lags, see: OECD (1976), *Methods of Accelerating the Collection and Circulation of R&D Data*, DSTI/SPR/76.52.
[67] OECD (1978), *op. cit*, pp. 16-17.
[68] OECD (1999), *Updating the STAN Industrial Database Using Short Term Indicators*, DSTI/EAS/IND/WP9 (94) 13. p. 3.
[69] OECD (1978), *op. cit*. p. 20.

University R&D

The data on university R&D have always been of very poor quality, so much so that an OECD draft report on trends in university R&D, intended for official publication, would never be published: [70] the data were qualified as "rather unsatisfactory" because of "serious conceptual and practical problems". [71] As John Irvine et al. noted in a subsequent study on academic R&D for the UK government, the statistics on university R&D "are of increasingly limited utility for policy purposes". [72] The problem had been a primary concern for the third OECD users group, [73] and the Member countries' experts themselves felt that:

> Some of the data used in the report (especially those for basic research) were not accurate enough to permit detailed conclusions to be drawn. (…) The major stumbling block was that a significant number of Member countries do not actually survey R&D in the Higher Education sector but rather make estimates by applying coefficients. (…) [Moreover] the preliminary draft was not sufficiently oriented towards problems of policy interest mainly because standard OECD surveys do not provide R&D data specifically relevant to these questions. [74]

The OECD thus accompanied the draft report on university R&D with a document identifying the main problems. [75] How, for example, could a country spend twice as much as another on university research and yet report similar numbers of university personnel affected to R&D? Why did expenditures on basic research differ by a ratio of 1 to 2 between otherwise similar countries? The answer was: [76]

[70] OECD (1979), *Trends in R&D in the Higher Education Sector in OECD Member Countries Since 1965 and Their Impact on National Basic Research Efforts*, SPT (79) 20.
[71] *Ibid*, p. 1.
[72] J. Irvine, B. Martin, and P. Isard, *Investing in the Future: An International Comparison of Government Funding of Academic and Related Research*, Worchester: Billing and Sons Ltd, p. 5.
[73] OECD (1985), *Report of the Third Ad Hoc Review on Science and Technology Indicators*, SPT (85) 3.
[74] OECD (1980), *Report of the Meeting of NESTI*, DSTI/SPR/80.7, pp. 2-3.
[75] OECD (1979), *National Methods of Measuring Higher Education R&D: Problems of International Comparison*, SPT (79) 21.
[76] OECD (1985), *Methods Used in OECD Member Countries to Measure the Amount of Resources to Research in Universities with Special Reference to the Use of Time-Budget Series*, DSTI/SPR/85.21, p. 1.

> Establishing R&D statistics for the higher education sector is particularly difficult compared with the situation for the other traditional sectors of R&D performance, as the accounting systems of the universities do not in general give any information about the amount of R&D financed by general university funds. These funds are given jointly to the activities performed in the universities: teaching, research and other activities. The producer of statistics has therefore to develop methods of identifying and measuring the R&D part separately from total activities in the universities.

The sources of the discrepancies are many, but for the present purposes it will suffice to mention only the major ones: [77] coverage of the university sector differed according to country (some institutions, like university hospitals and national research councils, were treated differently); estimates were used in place of surveys because they were cheaper, and coefficients derived from the estimates were little more than informed guesswork and were frequently out-of-date; general university funds were attributed either to the funder or to the performer; the level of aggregation (fields of science classification) was generally not detailed enough to warrant analysis; finally, there was a great deal of subjectivity involved in classifying research activities, according to a basic/applied scheme that was "no longer used in certain countries, although policy makers still persists in requesting such data in spite of its many shortcomings" [78] - the OECD itself no longer used the classification in its analyses. [79]

These difficulties led to a small-scale survey of national methods for measuring resources devoted to university research in 1981, [80] updated in 1983, [81] a workshop on the measurement of R&D in higher education in 1985 [82] and, as a follow-up, a Supplement to

[77] Some of these were already well identified in 1969. See: OECD (1969), *The Financing and Performance of Fundamental Research in the OECD Member Countries*, DAS/SPR/69.19, p. 4.

[78] OECD (1986), *Summary Record of the OECD Workshop on Science and Technology Indicators in the Higher Education Sector*, DSTI/SPR/85.60, p. 24.

[79] In its 1986 edition of *Recent Trends in Total R&D Resources*, the OECD did not include any data on basic research because of problems with the available data: OECD (1986), *Recent Trends in Total R&D Resources in OECD Member Countries*, DSTI/IP/86.05.

[80] OECD (1981), *Comparison of National Methods of Measuring Resources Devoted to University Research*, DSTI/SPR/81.44.

[81] OECD (1984), *Comparison of National Methods of Measuring Resources Devoted to University Research*, DSTI/SPR/83.14.

[82] *Op. cit*. note 73.

the Frascati manual in 1989, [83] which was later incorporated into the manual as Annex 3. The Supplement recommended norms for the coverage of the university sector, the activities and types of costs to be included in research, and the measurement of R&D personnel.

Despite the Supplement, however, higher education data continued to be "the least satisfactory in terms of quantity, quality, and details (…). Countries appear to be less motivated to spend additional resources on improving their submissions of Higher Education R&D expenditures (HERD) and manpower (HEMP) data than for their industrial R&D statistics". [84] In fact, industrial R&D accounted for the bulk (60%) of total R&D in most of the OECD countries, and was prioritized by the government policies of the time. [85] The OECD concluded: [86]

> There is relatively little that the OECD Secretariat itself can do to improve the comparability of the data that it issues, unless Member countries themselves make the necessary efforts (…).

Business R&D

Ameliorating the quality of industrial R&D data was the second challenge the STIU faced in its efforts to increase international comparability. Contrary to the unpublished paper on university R&D, a study of trends in industrial R&D was issued in 1979. [87] Few years before, the statistical unit had already identified the main problems with industrial R&D in

[83] OECD (1989), *The Measurement of Scientific and Technical Activities: R&D Statistics and Output Measurement in the Higher Education Sector*, Paris.
[84] OECD (1995), *The Measurement of University R&D: Principal Problems of International Comparability*, CCET/DSTI (95) 110.
[85] The same bias holds for the private non-profit sector where the relatively small size of R&D – around 5% - influenced the low priority given to it in statistics. See: K. Wille-Maus (1991), *Private Non-Profit Sector and Borderline Institutions*, DSTI/STII (91) 22, p. 5.; P. Jones (1996), *The Measurement of Private Non-Profit R&D: Practices in OECD Member Countries*, DSTI/EAS/STP/NESTI (96) 7.
[86] OECD (1986), *op. cit*, p. 24.
[87] OECD (1979), *Trends in Industrial R&D in Selected OECD Countries, 1967-1975*, Paris.

an internal document, [88] listing three broad classes of problems in comparability: comparisons across countries, time, and other economic variables.

The major difficulties were the following: differences, that were indeed very large, [89] between governmental and corporate estimates of public funds to industry, particularly in defense and aerospace; bad coverage of certain industries (agriculture, mining and services) on the assumption that they did not undertake much R&D, or because they were included in another sector (government); uneven treatment of the ways by which governments indirectly support industrial R&D (as in the form of loans or tax exemptions) which caused variations of over 9% in statistics; different classifications of R&D activities (principal industry or product field); changes in surveys over time as when over half of the increase in the number of firms performing R&D in France, was due to the broadening of the survey, or as in the cases in which statistics evolved *en dents de scie* from year to year (Japan); and the difficulty, if not impossibility, of comparing R&D with economic, demographic, and trade statistics. [90]

Despite these limitations, the OECD believed the data inspired <u>reasonable confidence</u>: [91] "Anyone who has worked with financial data on production, measures of labour force or on scientific activities is bound to acknowledge that R&D data are just <u>as reliable as</u> other existing economic and scientific statistics". [92] This was a variant of the 1950s' antirealist argument. The OECD nevertheless added: "The Secretariat finds itself in a situation of apparent conflict where it is hard to reconcile the daily demand for highly detailed and internationally comparable information with the actual production of data available which are often aggregated to a much greater degree and difficult to compare at a detailed level" (p. 16).

[88] OECD (1975), *Statistical Problems Posed by an Analysis of Industrial R&D in OECD Member Countries*, DSTI/SPR/75.77
[89] 400 million DM in Germany, 362 million francs in France, and 17 million pounds in United Kingdom.
[90] Other differences concerned the measurement of R&D personnel: the number of persons working on R&D varied from country to country according to whether it was measured for a given date or during a given period, and whether by occupation or qualification.
[91] OECD (1978), *Comparability and Reliability of R&D Data in the Business Sector*, DSTI/SPR/78.4, p. 4.
[92] *Ibid*. p. 3.

In the decade that followed, the OECD invested considerable efforts in data harmonization in order to link R&D statistics with economic ones, and to enlarge the STAN (Structural Analysis) database. [93] In fact, "the establishment of the STAN database has revealed not only theoretical problems of comparability but the fact that the current system no longer produces a set of data suitable for purely R&D analysis (…). With current data, there are very few detailed indicators for which reliable international comparisons can be made across a reasonable number of countries". [94] The work involved three actions: 1) developing compatible classifications (applying firm instead of an entire enterprise as the unit of analysis, and applying the product group rather than the principal activity as the unit of classification); 2) collecting data according to the standard industrial classification SIC while adding subclasses for high-intensive industries; and 3) distinguishing product and process R&D to better measure flows between industries. As a consequence, the OECD created a database called ANBERD (Analytical Business Enterprise R&D), which included many OECD estimates that differed from national data. [95]

Despite these efforts, huge methodological differences persisted between countries: [96] 1) to avoid overly expensive surveys, for example, countries usually considered only specific industries, measured only significant R&D activities, and excluded small companies. [97] And, depending on the country, these companies ranged in size from 5 to 100 employees since the Frascati manual never proposed a threshold; [98] 2) some countries performed estimates for non-respondents, but these varied considerably and were of irregular quality; [99] 3) questionnaires (and wording of questions) varied from country to country; [100] 4) above

[93] OECD (1990), *Propositions concernant le traitement des données détaillées sur la R&D industrielle*, DSTI/IP (90) 20.
[94] J.F. Minder (1991), *Treatment of Industrial R&D Data*, DSTI/STII (91) 17, p. 3.
[95] OECD (2001), *ANBERD Database*, NESTI Meeting, Rome 14-15 May.
[96] OECD (2000), *Examples of National Practices in R&D Surveys*, DSTI/EAS/STP/NESTI (2000) 19.
[97] See B. Godin (2000), Defining R&D, *op. cit*.
[98] For a summary of differences in coverage, population and method of surveys, see: M. Akerblom (2001), *op.cit*.
[99] The imputation methods used to estimate missing values varied. Some used information taken from the same survey, others from previous surveys or some related source, still others preferred statistical techniques.
[100] C. Grenzmann (2001), *Develop Standard Questionnaire Particularly for BE Sector*, DSTI/EAS/STP/NESTI (2001) 14/PART1.

all, "no country actually is using a pure statistical approach in surveying R&D activities", for example surveying all units or a statistical sample. [101] R&D surveys are in several countries censuses of enterprises known or supposed to perform R&D or a combination of a statistical census/sample approach and surveying known or supposed R&D performers approach based on lists of enterprises receiving government support or tax reductions for R&D. Taking into account information on previous or probable R&D therefore excludes a lot of enterprises, mainly SME.

Government R&D

The third and final problem to which the OECD devoted itself was the difficulties involved in comparing government R&D. [102] As with university, few countries even bothered to survey government R&D. [103] Rather, they usually relied on the functional allocation of the R&D appropriations that were reflected in the budgets, which amounted to identifying all budget items involving R&D.

The OECD soon recognized a problem with this practice: appropriation only measured plans and intentions, whereas expenditures measured the money that was actually spent. As a result, the amount of government R&D funding reported by performer and funder was never the same. A recent Norwegian study compared the results obtained from surveys with those estimated from budgets. [104] At the macro level, the two data sets gave roughly the same total Government R&D expenditures. The deviation was only 1 million on a total of 8,9 billion NOK. But large deviations were observed at the more detailed level in the case of Defense, Education, and Health. The main sources of discrepancies were the difficulty in interpreting the concept of development, and the different treatment of related scientific

[101] M. Akerblom (2001), *op. cit*., p. 7.
[102] OECD (1972), *The Problems of Comparing National Priorities for Government Funded R&D*, DAS/SPR/72.59. This document became chapter 2 of OECD (1975), *Changing Priorities for Government R&D*, Paris.
[103] Exceptions are Canada and the United Kingdom. Other countries either made text analysis of budgets or estimate appropriations from budget documents. For methodologies used in European countries, see: Eurostat (1995), *Government R&D Appropriations: General University Funds*, DSTI/STP/NESTI/SUR (95) 3, pp. 2-3.
[104] O. Wiig (2000), *Problems in the Measurement of Government Budget Appropriations or Outlays for R&D (GBAORD)*, DSTI/EAS/STP/NESTI (2000) 25.

activities (RSA) such as policy studies and evaluation. In the United States, another evaluation estimated the difference at around 30%. [105] The gap was so huge ($5 billion) that the US Senate Committee on Commerce, Science, and Transportation asked the General Accounting Office (GAO) to review the procedure because of concerns over whether Members of Congress could truly rely on the NSF's data. [106] Consequently, at the suggestion of the United States, [107] the OECD recently decided to include a paragraph in the next edition of the Frascati manual recognizing "the likelihood of differences in R&D expenditure totals between those estimated from the funders and those estimated from the performers of R&D. However, not too much emphasis should be put on this, as it might raise excessive suspicion on published R&D data (performer based)". [108]

To better harmonize national practices, a draft supplement to the Frascati manual specifically devoted to the measurement of government-funded R&D was completed in 1978. [109] It dealt with the information available in budgetary documents, the coverage of the sector, the type of funds to include, and the classification by socioeconomic objectives. The draft manual was never published as a separate publication. In fact, these data "play only a modest role in the general battery of S&T indicators and do not merit a separate manual" stated the OECD. [110] Instead of a manual, the specifications were relegated in an abridged form to a chapter in the fourth edition of the manual. [111]

[105] NSF (1999), *Study on Federally Funded Industrial R&D: Summary of Finding from Company Interviews and Analyses of Collateral Data*, DSTI/EAS/STP/NESTI (99) 2; OECD (2001), *Reconciling Performer and Funder R&D*, DSTI/EAS/STP/NESTI (2001) 13.
[106] GAO (2001), *R&D Funding: Reported Gap Between Data From Federal Agencies and Their R&D Performers Results From Noncomparable Data*, GAO-01-512R, Washington; M.E. Davey and R.E. Rowberg (2000), *Challenges in Collecting and Reporting Federal R&D Data*, Washington: Congressional Research Service.
[107] J.E. Jankowski (2001), *Relationship Between Data from R&D Funders and Performers*, DSTI/EAS/STP/NESTI (2001) 14/PART7.
[108] OECD (2001), Summary of the Main Conclusions of the Meeting on the Revision of the Frascati Manual held 9-11 May 2001, Annex to OECD (2001), *Summary Record*, DSTI/EAS/STP/NESTI/M (2001) 1, p. 15.
[109] OECD (1978), *Draft Guidelines for Reporting Government R&D Funding by Socio-Economic Objectives: Proposed Supplement to the Frascati Manual*, DSTI/SPR/78.40.
[110] OECD (1991), *Classification by Socio-Economic Objectives*, DSTI/STII (91) 19, p. 9.
[111] For the problems specifically related to the classification of government-funded R&D, see: B. Godin (2001), *Tradition and Innovation: The Historical Contingency of R&D Statistical Classifications*, Montreal: OST.

In the nineties, the OECD documented two more discrepancies in the data. Firstly, and following the "Subsidies and Structural Adjustment" project started in 1987, [112] the Industry Committee reported how statistical analyses of government-funded R&D were biased by the failure to appropriately measure indirect governmental support of industrial R&D, such as tax incentives. [113] The work conducted to a database on 1 000 public support programs (of which nearly 300 covers R&D and technological innovation), a manual on industrial subsidies published in 1995, [114] and, as regard science and technology, a study on national technology programs. [115] Few years later, during the *Technology, Productivity and Job Creation* project of the Directorate for Science, Technology and Industry (DSTI), the statistical division measured the impact of past practices on the data, since the sums (of fiscal measures) were not imputed to government R&D, [116] as recommended by the Frascati manual (1993). The biases were particularly significant in the case of the United States, Canada and Australia. To include these types of government R&D support, however, would entail extending the R&D survey to post-R&D activities and related scientific activities, and would "probably mean drafting a separate manual". [117]

The second discrepancy concerned the increasing internationalization of R&D activities, which resulted in an incomplete picture of public R&D funding. European countries, for instance, included neither the estimated R&D content of their contribution to the European Community budget, nor their receipts from abroad as government R&D. Although the latter was of little statistical consequence in countries with large R&D efforts, its effects were

[112] OECD (1992), *Subsidies and Structural Adjustment: Draft Report to the Council at Ministerial Level*, DSTI/IND (92) 8.
[113] OECD (1990), *The Industry's Committee on "Subsidies and Structural Adjustment" Project and Its Probable Implications for the Frascati Manual*, DSTI/IP (90) 24; T. Hatzichronoglou (1991), *The Measurement of Government R&D Funding in the Business Enterprise Sector*, DSTI/STII (91) 30.
[114] OECD (1995), *Industrial Subsidies: A Reporting Manual*, Paris; OECD (1997), *OECD Sources of Data on Government Support for Industrial Technology: Coverage, Availability and Problems of Compilation and Comparison*, DSTI/IND/STP/SUB/NESTI (97) 1; OECD (1997), *Thematic Analysis of Public Support to Industrial R&D Efforts*, DSTI/IND/SUB (97) 13/REV 1.
[115] OECD (1993), *The Impacts of National Technology Programs*, DSTI/STP (93) 3. Officially published in 1995.
[116] OECD (1998*), Measuring Government Support for Industrial Technology*, DSTI/EAS/STP/NESTI (98) 11.
[117] A. Young (1999), *Some Lessons of the Study on Government Support for Industrial Technology for Future Editions of the Frascati and Oslo Manuals*, DSTI/EAS/STP/NESTI (99) 4, p. 7.

much more strongly felt, in fact twice as much for Greece and Ireland, in small R&D intensive countries. [118]

These two problems – governmental support of industrial R&D and internationalization - would lead to new methodological specifications (see below). As was the case with other sectors then, problems remained despite normalization. To take one more example, only central government appropriations were included in the United States, Switzerland, and Sweden because the amount of R&D supported by local authorities is thought to be negligible.

Metadata

The persistence of differences in national practices led the OECD to start a regular series entitled *Sources and Methods*. Each sector (government, business, higher education) was the object of detailed technical notes, which were either included in the corresponding statistical publication [119] or published separately in greater details – other topics have been recently added (Table 1). [120]

Sources and Methods are technical documents containing methodological notes from Member countries and the OECD: "Countries are requested to send a set of methodological notes with their responses to OECD R&D surveys describing the situation in the survey year, any changes since the previous survey and differences from the specifications in the Frascati manual". [121] The information so communicated to the OECD constitutes the bulk of *Sources and Methods* and are called metadata. The OECD defines metadata as "the information associated with a statistical point or series which contains the information

[118] OECD (1997), *Treatment of European Commission Funds in R&D Surveys: Summary of National Practices*, DSTI/EAS/STP/NESTI/RD (97) 3; OECD (1998), *Measuring the Internationalization of Government Funding of R&D*, DSTI/EAS/STP/NESTI (98) 3.
[119] See documents listed in Annex 3 and 4.
[120] A Database is actually in progress to integrate Sources and Methods for each sector. See: OECD (2001), *OECD R&D Sources and Methods Database*, DSTI/EAS/STP/NESTI (2001) 13; OECD (2000), *OECD Access Database on R&D Sources and Methods*, DSTI/EAS/STP/NESTI (2000) 32.
[121] OECD (1994), *Metadata, Sources and Methods*, DSTI/EAS/STP/NESTI (94) 12, p. 3.

needed by a person who has no first hand experience of the survey underlying the data but who wished to interpret and analyze it and compare it with points or series from other sources, for example from other countries or taken at other points in time". [122]

Table 1.
Sources and Methods

Government R&D

Socioeconomic Objectives (1982, 1987, 1990) [123]
International Funding (1998) [124]
Industrial Support (1998) [125]

Business R&D (1992, 1997) [126]

Higher Education R&D (1997) [127]

R&D - General (2000) [128]

Technology Balance of Payments (2000) [129]

The structure of *Sources and Methods* is quite simple: for each country, arranged alphabetically, the major points of departure from the OECD standards are presented and explained. These notes are far more detailed that the standard footnotes that appeared in

[122] OECD (1994), *op. cit,* p. 2.
[123] OECD (1982), *Sources and Methods (Volume A) : The Objectives of Government R&D Funding*, DSTI/SPR/82.06; OECD (1987), *Sources and Methods (Volume A) : The Objectives of Government Budget Appropriations of Outlays for R&D*, DSTI/IP/87.13
[124] OECD (1998), *Measuring the Internationalization of Government Funding: Sources and Methods*, DSTI/EAS/STP/NESTI/RD (98) 2
[125] OECD (1998), *Measuring Government Funding of Industrial Technology: Sources and Methods*, DSTI/EAS/STP/NESTI/RD (98) 4
[126] OECD (1992), *Dépenses de R-D du secteur des entreprises dans les pays de l'OCDE: données au niveau détaillé des branches industrielles de 1973 à 1990*, OECD/GD (92) 173 ; OECD (1998), *R&D in Industry: Expenditures and Researchers, Scientists and Engineers - Sources and Methods*, Paris.
[127] OECD (1997), *Measuring R&D in the Higher Education Sector: Methods Used in the OECD/EU Member Countries*, DSTI/EAS/STP/NESTI (97) 2.
[128] OECD (2000), *R&D Statistics: Sources and Methods*, Paris.
[129] OECD (2000), *Technology Balance of Payments: Sources and Methods*, Paris.

statistical repertories like *Main Science and Technology Indicators*, where only the essential differences between countries are described (Annex 6). They deal with the coverage of the sector, the source of the data, the classification used, and the period and years available.

A brief scan of the metadata is enough to shake one's confidence in the reliability of international R&D statistics. In fact, standard footnotes do not give the whole story. One has to look at *Sources and Methods* to really appreciate the OECD statistics. To take but a few examples, the following standard footnotes are associated to the following national statistics (see Annex 4):

France	Provisional
Germany	Estimate or projection
Japan	Break in series
	Excludes GUP
Sweden	Break in series
	Underestimate
United Kingdom	None
United States	Excludes capital expenditures
	Provisional

However, countries' methodological notes indicate the more complete following limitations. [130] *France*: the *Centre National de la Recherche Scientifique* (CNRS) is included in the higher education sector whereas in other countries, such as Italy, this type of organization is classified in the government sector. As from 1997, the method used to evaluate R&D personnel has changed. As from 1992, data for enterprise and government sectors are not comparable to the 1991 counterparts. *Germany*: current data cannot be compared with pre-1991 data because of a statistical break caused by reunification. The recent inclusion in German statistics, moreover, of graduate students as active researchers

[130] Taken from: OECD (1999), *Basic Science and Technology Statistics*, Paris.

and grant recipients constitutes another statistical break. As from 1993, R&D expenditures in the government sector include R&D performed abroad. *Japan*: there is an important overestimation of R&D personnel in Japan, by about 25% (and of GERD by about 15%), because the data are not expressed in full-time equivalents. The data on R&D by socio-economic objectives are underestimated because military contracts are excluded. *United Kingdom*: business R&D includes funds that accrue from other sectors. Due to privatizations, several organizations have changed sectors causing changes in statistics broken down by sectors. Hospitals are included in the government sector. *United States*: there was a break in its data series in 1991 because of a change in the method of allocating institutional funds to universities and of the exclusion of capital expenditures in the university sector. These modifications reduced the US government's contribution by around 20-25%. The United States only includes the central government in its calculation of government R&D, and the social sciences and humanities are generally excluded from total R&D.

Taken one by one, these limitations had negligible effects on the overall statistical results. When considered together, however, they summed up to quite large variations. Overall, one must accept using numbers "allowing a margin of error of a few billions [dollars]". [131] As the first OECD experimental study linking R&D data to economic ones explained: [132] "in the present situation it is not possible to assess the accumulated global impact of all these separate distortions (…). If taken individually, the conceptual methodological or actual distortions between the R&D and non-R&D data in most cases have probably no very serious effect, but it is impossible to venture even a rough approximation of the cumulative impact of all of them". Since that time, however, some authors have produced such estimates. [133]

While the technical notes which are indispensable for interpreting the statistics were usually systematically included in the OECD statistical series, very few of them were included in

[131] D.S. Greenberg (2001), *Science. Money, and Politics: Political Triumph and Ethical Erosion*, Chicago: University of Chicago Press, p. 79.
[132] OECD (1976), *Comparing R&D Data with Economic and Manpower Series*, DSTI/SPR/76.45, pp. 6-7.

the science and technology policy reports – a questionable omission considering, as we have seen, that the OECD had on occasion decided not to publish whole analyses (like that on university R&D) or specific numbers (on basic research for example) because of the poor quality of the data. Whether one examines the *Science and Technology Indicators* series of the eighties or current editions of *Science, Technology and Industry Outlook*, one would be hard pressed to find interpretative details or even rudimentary qualifications. The construction of statistics remains a black box, and the consequences this has for policy remain equally obscure. Since there is no way of evaluating the optimal level of resources a country should invest in science and technology, governments must compare their performance to that of other countries in planning their policies, which are liable to move along the wrong track when important caveats are overlooked.

But the sheer quantity of notes has now grown so large that national experts (NESTI) are beginning to downplay their importance and to suggest retaining "only those that are essential for the understanding of the data (mainly notes showing a break in series or a methodological change, an overestimation or an underestimation of the data)". [134] A curious proposition since it was the same class of experts who had first requested the notes over twenty years ago. [135]

Conclusion

Constructing statistics is no easy task, and using them is an art. Before the sixties, there were wide differences in national statistics depending on the agency conducting the survey: definitions, coverage, demarcations, samples and estimates were all different. As soon as one became interested in comparing countries, new problems of comparability began appearing: national measures were of unequal quality, each country had its own specificities (like the organization of the university sector), and the national practices of

[133] Anderson et al. (1992),

[134] OECD (1998), *How to Improve the Main Science and Technology Indicators: First Suggestions From Users*, DSTI/EAS/STP/NESTI/RD (98) 9.

[135] The absence of methodological notes and analysis of limitations was one of the criticisms of NSF's *Science Indicators* in the 1970s. See: B. Godin (2001), *The Emergence of Science and Technology Indicators: Why Did Governments Supplement Statistics with Indicators*, Montreal: OST.

statistical agencies were varying (year of survey, for example). As if these problems were not enough, a strange paradox was soon to appear: statistics from national R&D surveys did not correspond to those produced by innovation surveys. [136] Innovation surveys indicated significantly more R&D activity than did standard R&D surveys because of methodological differences that had been known since the eighties. [137] To obtain singular figures entailed, for many, combining the two surveys or at least having them conducted by the same agency. [138]

Over the last forty years, the OECD has advanced several solutions for harmonizing R&D statistics. It first developed international standards in 1963, set up a system of footnotes in the 1980s, then calculated estimates for missing national data and brought continual clarifications to its international norms (the Frascati manual is now in its fifth edition). These moves were accompanied with, or resulted in two consequences. Firstly, the construction of statistics eliminated national specificities. Normalization suppressed national differences in order to better compare countries. Secondly, an antirealist rhetoric was offered that minimized the limitations. While R&D statistics were increasingly used to support science policy analyses, [139] to study the impact of government funding of R&D [140] and to understand productivity issues, [141] the data's limitations were minimized or rarely discussed in published analytical documents. [142]

[136] OECD (2001), *Assess Whether There Are Changes Needed as a Result of the Comparison of R&D Data Collected in R&D and in Innovations Surveys*, DSTI/EAS/STP/NESTI (2001) 14/PART3; OECD (2000), *Measuring R&D in R&D and Innovation Surveys: Analysis of Causes of Divergence in Nine OECD Countries*, DSTI/EAS/STP/NESTI (2000) 26; OECD (1999), *Old and New Paradigms in the Measurement of R&D*, DSTI/EAS/STP/NESTI (99) 13.
[137] DSTI/IP/88.27
[138] Eurostat (2001), *Working Party Meeting on R&D and Innovation Statistics: Main Conclusions*, 19-20 April.
[139] See the 1980s series *Science and Technology Indicators* and the 1990s series *Science, Technology and Industry Outlook*.
[140] OECD (2000), The Impact of Public R&D Expenditures on Business R&D, in *Science, Technology and Industry Outlook*, pp. 185-200.
[141] OECD (2000), *R&D and Productivity Growth: A Panel Data Analysis of 16 OECD Countries*, DSTI/EAS/STP/NESTI (2000) 40. The last meeting of NESTI admitted that the study had numerous shortcomings and was rather mechanical. See: OECD (2001), *Summary Record*, DSTI/EAS/STP/NESTI (2001) 1, p. 7.
[142] For a similar criticism concerning NSF's *Science and Engineering Indicators*, see: B. Godin (2000), *The Emergence of Science and Technology Indicators: Why Did Governments Supplement Statistics With Indicators*, Montreal :OST.

C. Silver, chairman of the OECD first users group on R&D statistics, wrote, in the introductory remarks to its report: [143]

> I started my task as a skeptic and completed it converted – converted that is, to the view that policy makers use and even depend on R&D statistics and particularly on those giving comparisons of national efforts in particular fields. What I beg leave to question now is whether perhaps too much reliance is placed on these all-too-fallible statistics.

He was right. R&D statistics, like many social or economic statistics, are fallible numbers. Contrary to what one would expect, however, it is precisely the exercise of improving the statistics themselves that makes their reliability at times questionable: surveys (and their scope) are constantly improved because of changes in the evolution of concepts, statistical frameworks and systems of national accounting; because of institutional restructuring, scientific and technological advances, and the emergence of new policy priorities; and, above all, because of the knowledge gained from experiences and revision work. [144] It is these (positive) changes that often render statistics poorly comparable over time.

Comparisons in space, that is between countries, also have their limitations. Again, contrary to what one would expect, [145] the adoption of international standards did not eliminate differences. Although the Frascati manual "has probably been one of the most influential documents issued by this Directorate over the fifteen years of its existence", [146] and although the Science and Technology Indicators Unit (STIU) was "the only comprehensive source of reliable and internationally comparable S&T statistics", [147] international comparisons remain extremely difficult to do, if not altogether impossible, given the

[143] OECD (1973), *Report of the Ad Hoc Review Group on R&D Statistics*, SPT (73) 14, p. 6.
[144] OECD (1994), *Report on the Conference on Science and Technology Indicators in Central and Eastern European Countries*, CCET/DSTI/EAS (94) 12, p. 8.
[145] The first edition of the Frascati manual suggested that national "variations may be gradually reduced" with standardization (p. 6).
[146] OECD (1979), Notes by the Secretariat on the Revision of the Frascati Manual, DSTI/SPR/79.37, p. iii.
[147] OECD (1985), Report of the Third Ad Hoc Review, *op. cit*. p. 7.

differences in national specificities and practices. Yet avoiding international (or temporal) comparisons is virtually unthinkable, since statistics are meaningless without them. [148]

[148] S. Patriarca (1996), *Numbers and Nationhood: Writing Statistics in 19th Century Italy*, Cambridge: Cambridge University Press, p. 165.

Annex 1.
Coverage of Activities in Early R&D Surveys

	Inclusions	Exclusions
Government R&D		
Nat. Res. Committee (1938)	collection of data	routine work
Kilgore (1945)	pilot plant	social sciences routine exploration design
Canadian DRS (1947)	surveys analysis administration dissemination	routine work
NSF (1953)	social sciences indirect costs	routine work mapping and surveys exploration dissemination training
Dominion Bureau (1960)	planning and administration capital data collection scientific information scholarship and fellowship	
Industrial R&D		
FBI (1947)		routine work testing
Harvard Bus. School (1953)	scale, pilot plant design information	exploration market research economic research legal work
Dominion Bureau (1955)		market research routine patent work advertising social sciences exploration
NSF (1956)	pilot plant design laboratory scale prototypes	market research economic studies legal work technical services
DSIR (1958)	design prototypes	routine work tooling up market research

Annex 2.
Differences in Definition and Coverage
According to Freeman and Young

Expenditures

a. Not all countries included the social sciences in the survey (UK), and when they did it did not cover the enterprise sector.
b. Some countries included depreciation of capital expenditures (United States, Germany),
c. Some countries (France) included contributions to international research organizations in their national statistics,
d. Only Norway estimated related scientific activities in order to exclude them from R&D,
e. The French derived R&D expenditures funded by government from the source of funds rather than from the performer,
f. High rate of non-responses from industry (France),

Manpower

g. Varying definitions of scientists and engineers (some are based on qualifications, others on occupations)
h. Only a few countries (United States) estimated full-time equivalent personnel for the higher education sector,
i. A more liberal definition of support personnel in the United States than in Europe.

Others

j. Varying definition of sectors according to countries,
k. Difficulties in estimating the basic/applied dimensions of research,
l. Classification of enterprises by industries rather than by products,
m. Exchange rates missing.

Annex 3.

OECD Standard Footnotes [149]

a. Break in series with previous year for which data is available
b. Secretariat estimate or projection based on national sources
c. National estimate or projection adjusted, if necessary, by the Secretariat to meet OECD norms
d. (Note used only for internal OECD data-processing)
e. National results adjusted by the Secretariat to meet OECD norms
f. Including R&D in the social sciences and humanities
g. Excluding the social sciences and humanities
h. Federal or central government only
i. Excludes data from the R&D content of general payment to the higher education sector for combined education and research (public GUF)
j. Excludes most or all capital expenditures
k. Total intramural R&D expenditures instead of current intramural R&D expenditures
l. Overestimated or based on overestimated data
m. Underestimated or based on underestimated data
n. Included elsewhere
o. Includes other classes
p. Provisional
q. At current exchange rate and not at current purchasing power parities
r. Including international patents applications.
s. Unrevised breakdown not adding to the revised total
t. Do not correspond exactly to the OECD recommendations
u. Includes extramural R&D expenditures
v. The sum of the breakdown does not add to the total

[149] OECD (2000), *Main Science and Technology Indicators (2)*, Paris.

Annex 4.

GERD (million current PPP$)

	1993	1994	1995	1996	1997	1998	1999
Australia		5 581.2		6 776.3		6 758.7	
Austria	2 286.0	2 511.4[c]	2 686.7[c]	2 895.0[c]	3 123.6	3 476.8[c]	3 644.7[c, p]
Belgium[p]	3 491.3[a]	3 645.2	3 853.1	4 105.7	4 270.6		
Canada	9 188.4	10 110.9	11 051.4	11 059.1	11 711.2	12 366.9[c]	12 775.0[p]
Czech Republic	1 302.2[t]	1 249.5[t]	1 293.3[a]	1 391.3	1 569.8	1 680.3	1 721.7
Denmark	1 786.4		2 203.1	2 360.4[c]	2 534.0	2 604.1[c]	2 770.0[c]
Finland	1 754.3	1 943.3	2 203.6	2 529.1[c]	2 858.6	3 246.5	3 652.4[c, p]
France	26 441.6	26 520.1	27 722.6	27 783.8	27 060.8	27 880.4[p]	
Germany	36 186.6	37 028.5[c]	39 451.5[c]	39 902.3[c]	41 751.5[c]	43 556.7[c]	46 218.0[c]
Greece	545.0		652.0		721.7		
Hungary	802.5[t]	759.3[a]	680.4	618.7	721.4	720.3	763.9
Iceland	66.3	72.0	91.7		120.4	141.1[c]	137.5[c]
Ireland[c]	609.8	749.7	877.2	947.3	1 083.8		
Italy	11 483.2	11 343.7	11 522.8	12 100.8	11 913.4	12 613.2[p]	13 310.7[p]
Japan	74 382.2[i]	75 287.2[i]	84 783.3[i]	85 469.6[a]	89 632.5	92 663.1	
Korea	10 403.2	12 771.4	15 345.7	17 287.5	19 000.0	16 980.7	
Mexico	1 303.0[c]	1 831.4	1 923.1	2 066.0	2 441.8		
Netherlands	5 456.8	5 880.0[a]	6 528.9	6 837.8	7 378.0	7 391.7	
New Zealand	545.6		606.2		752.1		
Norway	1 597.0		1 739.6[a]		1 951.8		2 145.6[p]
Poland		1 720.5	1 875.6	2 023.5	2 010.3	2 258.8	2 456.6
Portugal			774.5		946.3		
Spain	4 765.7	4 519.9[c]	4 838.6	5 182.9[c]	5 297.7	6 116.8[c]	6 443.5[c]
Sweden	4 984.0[a, m]		6 095.4[a, m]		6 845.4		
Switzerland				4 867.6			
Turkey	1 465.2	1 156.7	1 321.3	1 699.0	1 996.9		
United Kingdom	21 246.1	21 765.1	21 672.5	22 467.8	22 682.2	23 445.2	
U.S.A.[j]	165 868.0	169 270.0	183 694.0	196 995.0	212 246.0	226 653.0[p]	243 548.0[p]
Total OECD[b]	390 699.8	402 062.9	440 887.7[a]	468 026.1	494 388.3	518 113.7[p]	
North America	176 359.4	181 212.3	196 668.6	210 120.0	226 399.0	241 869.2[b, p]	259 315.3[b, p]
European Union	121 724.6[b]	124 629.5[b]	131 081.7	135 165.1[b]	138 490.7	144 989.7[b, p]	
Nordic countries	10 187.9[a]		12 333.5[a]		14 310.1		

Annex 5.
OECD Analytical Reports on R&D Data

1967 The Overall Level and Structure of R&D Efforts in OECD Member Countries
1971 R&D in OECD Member Countries: Trends and Objectives.
1975 Patterns of Resources Devoted to R&D in the OECD Area, 1963-1971.
1975 Changing Priorities for Government R&D: An Experimental Study of Trends in the Objectives of Government R&D Funding in 12 OECD Member Countries, 1961-1972.
1979 Trends in Industrial R&D in Selected OECD Countries, 1967-1975.
1979 Trends in R&D in the Higher Education Sector in OECD Member Countries Since 1965 and Their Impact on National Basic Research Efforts
1984 Science and Technology Indicators (1)
1986 Science and Technology Indicators (2)
1989 Science and Technology Indicators (3)

Annex 6.
Repertoires of Statistics (OECD)

1. International Survey of the Resources Devoted to R&D by OECD Members Countries (1967-83; biennial)
    1. 1967 to 1973 Four publications for each survey (one by sector and one general)
    2. 1975 to 1983 Fascicules by country (+ International Volume for one year only)
2. "Recent Results" and "Basic Statistical Series" (1980-83). The two documents would give the next two publications:
3. Main Science and Technology Indicators (1988-Today: twice a year)
4. Basic science and Technology Statistics (1991, 1997, 2000)
5. Research and Development Expenditure in Industry (1995, 1996, 1997, 1999)
6. Science, Technology and Industry Scoreboard of Indicators (1995, 1997, 1999)