



OECD Science, Technology and Industry Working Papers  
1997/01

Bibliometric Indicators  
and Analysis of Research  
Systems: Methods and  
Examples

**Yoshiko Okubo**

<https://dx.doi.org/10.1787/208277770603>

**STI WORKING PAPERS  
1997/1**

**BIBLIOMETRIC INDICATORS AND ANALYSIS OF RESEARCH SYSTEMS:  
METHODS AND EXAMPLES**

**Yoshiko Okubo**

**ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT**

**Paris**

**51765**

**Document complet disponible sur OLIS dans son format d'origine  
Complete document available on OLIS in its original format**

## **STI Working Papers Series**

This Working Papers Series of the OECD Directorate for Science, Technology and Industry is designed to make available to a wider readership selected studies prepared by staff in the Directorate or by outside consultants working on OECD projects. The papers included in the series are of a technical and analytical nature and deal with issues of data, methodology and empirical analysis in the areas of work of DSTI. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

Comment on the papers is invited, and should be sent to Directorate for Science, Technology and Industry, OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.

The opinions expressed in these papers are the full responsibility of the author(s) and do not necessarily reflect those of the OECD or of the governments of its Member countries.

**Copyright OECD, 1997**

**Application for permission to reproduce or translate all or part of this publication should be made to: Head of Publications Service, OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.**

## **Bibliometric indicators and analysis of research systems: Methods and examples**

Yoshiko Okubo\*

This report – linked to the technical documents of the OECD manuals for the measurement of R&D activities (“Frascati Family”) – presents the essential elements of bibliometrics and its application to the analysis of research systems. Bibliometrics is based on the enumeration and statistical analysis of scientific output in the form of articles, publications, citations, patents and other, more complex indicators. It is an important tool in evaluating research activities, laboratories and scientists, as well as the scientific specialisations and performance of countries. The report, having set the background for the development of bibliometrics, presents the databases on which bibliometrics is built, as well as the principal indicators used. Twenty-five examples are presented at the end of the document, illustrating the various uses of bibliometric methods for analysing research systems. These indicators measure scientific output, by counting the number of papers; the impact of papers on scientific disciplines, by counting the number of co-citations; the extent of international co-operation, as evidenced by joint signatures; the scientific content of patents, etc.

\* The original version of this paper was written in French by Yoshiko Okubo, Laboratoire Stratégie & Technologie, École Centrale Paris, Grande voie des Vignes, 92295 Châtenay-Malabry Cedex, France. Telephone: +33 [0]1.41.13.10.24. Fax: +33 [0]1.46.83.99.17. E-mail: yoshiko@cti.ecp.fr. This English-language version is a translation by the OECD.

The author would like to thank the following people for their help in gathering data, and for their criticism: Jennifer Bond, Tibor Braun, L. Grauls, Shlomo Herskovic, Terttu Luukkonen, Aris Kaloudis, Maurits Pino, Rosa Sancho, Per O. Seglen, Gunnar Sivertsen, Henry Small, Anthony Van Raan, Jan C.G. Van Steen and Gunnar Westholm, as well as Jean-Éric Aubert of the OECD Secretariat, for the preparation of this report in view of its publication.

## TABLE DES MATIERES

INTRODUCTION .....	6
CHAPTER 1. WHAT IS BIBLIOMETRICS?.....	8
CHAPTER 2. THE ADVENT OF BIBLIOMETRICS .....	10
Background.....	10
The need to evaluate research .....	11
Bibliometrics and the measurement of science.....	12
CHAPTER 3. BIBLIOMETRIC DATABASES.....	14
The main bibliometric databases.....	14
Problems posed by databases in general.....	15
Types of literature .....	15
The <i>Science Citation Index</i> (SCI) database .....	16
Structure and potential of the SCI database .....	16
The Science Citation Index and its limitations.....	17
Why do authors use citations? .....	17
“Negative citations” .....	17
The “uncited” .....	18
“Self-citation” .....	18
The language factor .....	18
The breakdown by scientific discipline .....	18
CHAPTER 4. USING BIBLIOMETRIC INDICATORS AND THE PRECAUTIONS TO TAKE .....	20
Introduction.....	20
The problems of co-authorship – “whole counting” or “fractional counting”? .....	21
The problem of database coverage .....	22
The marketing of bibliometric data and analysis .....	23
What does the future hold for bibliometric indicators?.....	23
CHAPTER 5. THE MAIN BIBLIOMETRIC INDICATORS AND THEIR APPLICATIONS .....	24
Introduction.....	24
Indicators of science and technology activity .....	24
The number of papers (Examples 1-4).....	24
The number of citations (Examples 5 and 6).....	25
The number of co-signers.....	25
The number of patents (Examples 10 and 11).....	26
The number of patent citations (Example 12) .....	27
Relational indicators (Examples 13 to 16).....	28
Co-publications.....	28

The affinity index (Example 17).....	28
Scientific links measured by citations (Examples 18 and 19).....	29
Correlations between scientific papers and patents (Examples 20-22).....	29
Co-citations (Example 23).....	29
The “co-occurrence” of words (Example 24) .....	30
Visual representation techniques for scientific fields and countries (Examples 25, 25A and 25B).....	30
<b>LIST OF EXAMPLES .....</b>	<b>32</b>
<b>NOTES.....</b>	<b>61</b>
<b>REFERENCES .....</b>	<b>62</b>
<b>ADDITIONAL REFERENCES FOR FURTHER READING .....</b>	<b>66</b>
<b>ANNEX I. INDEX OF ISO CODES .....</b>	<b>69</b>

## INTRODUCTION

Formerly absent from the concerns of most politicians, indicators of scientific activity are now at the heart of the debate over the linkages between advances in science and technology and economic and social progress. There is a growing awareness of the advantages of basing opinions, and subsequent choices, on criteria that lend themselves more to quantitative evaluation. Science policy reviews would seem inconceivable today without recourse to existing indicators. Long focused on measures of input, such as expenditure and R&D staff, interest is turning increasingly to output – and especially to technology-related output (*e.g.* patents, technology balance of payments, trade in high tech). Where science is concerned, bibliometric indicators are a must.

In this paper, we shall begin with a brief history of bibliometrics, to trace its origins and define its role in the evaluation of science. We shall present a variety of bibliometric indicators, noting their uses and their limitations, along with practical examples. The aim of this study is to highlight both the strengths and weaknesses of bibliometric indicators and, above all, the precautions that need to be taken when using them. It does not purport to be a comprehensive survey of the field.

Bibliometrics is a tool by which the state of science and technology can be observed through the overall production of scientific literature, at a given level of specialisation. It is a means for situating a country in relation to the world, an institution in relation to a country, and even individual scientists in relation to their own communities. These scientific indicators are equally suitable – providing the customary precautions are taken – for “macro” analysis (*e.g.* a given country’s share in global output of scientific literature over a specified period) and “micro” studies (*e.g.* a given institute’s role in producing papers in a highly circumscribed field of science). They constitute a way to assess the current state of science, which can help shed light on its structure. By providing new information, bibliometrics can be an aid to decision-making and research management. It alone cannot, of course, justify a decision or replace experts. Bibliometric indicators are practical tools which can be used in combination with other indicators.

The products of science are not objects but ideas, means of communication and reactions to the ideas of others. While it is possible simultaneously to track scientists and money invested, it is far more difficult to measure science as a body of ideas, or to grasp its interface with the economic and social system. For now, indicators remain essentially a unit of measure based on observations of science and technology as a system of activities rather than as a body of specific knowledge (National Science Foundation, 1989).

Each indicator has its advantages and its limitations, and care must be taken not to consider them as “absolute” indices. The “convergence” of indicators has to be tested in order to put the information they convey into perspective (Martin and Irvine, 1985).

It is a fact that there is a growing demand for bibliometric data from the scientific community itself. Most industrial countries publish sets of indicators similar to those of the National Science

Foundation, attesting to a perceived need. Faced with the task of convincing a parliament, a board of directors or the public at large – if not scientists and researchers themselves – it is essential to advance the right arguments, to be in full possession of the facts, and to back them up with “objective” figures.

## CHAPTER 1. WHAT IS BIBLIOMETRICS?

In 1969, Pritchard coined a new term – “bibliometrics” – for a type of study that had been in existence for half a century. The fact that Pritchard felt the need to redefine the scope of an area hitherto covered for fifty years by the term “statistical bibliography” (Hulme, 1923) demonstrated that a new field of quantitative research had emerged. For Pritchard, bibliometrics was defined as the application of mathematical and statistical methods to books and other means of communication (Pritchard, 1969, pp. 348-349).

Bibliometrics has become a generic term for a whole range of specific measurements and indicators; its purpose is to measure the output of scientific and technological research through data derived not only from scientific literature but from patents as well.

Bibliometric approaches, whereby science can be portrayed through the results obtained, are based on the notion that the essence of scientific research is the production of “knowledge” and that scientific literature is the constituent manifestation of that knowledge. Patents indicate a transfer of knowledge to industrial innovation and a transformation into something of commercial and social value; for this reason, they constitute an indicator of the tangible benefits of an intellectual and economic investment.

The idea that to publish their work (see the discussion below on what is considered a published work in bibliometrics) is the paramount activity of scientists has long been contented by science analysts. According to Price, a scientist is “...any person who has ever published a scientific paper” (Price, 1963). “Whenever a man labors, produces something new and the result is a publication, then he has been doing what I call science” (Price, 1969). His catchphrase “publish or perish” would suggest that publication of research findings is at the forefront of scientists’ activities.

To publish the results of their research is an obligation that scientists are compelled to fulfil (Merton, 1957b). New knowledge, updated by researchers, has to be transformed into information made available to the scientific community. Not only do scientists have to make their work available to the public at large, but they in turn are supposed to have access to the work of their peers. Research is carried out in a context of “exchange”. Even so, the fact that the system of scientific publication has survived in modern science is due, paradoxically, to scientists’ desire to protect their intellectual property. New scientific knowledge is a researcher’s personal creation, and claim to its discovery can be laid only through publication (Merton, 1957a).

The “reward system”, based on the recognition of work, merely underscores the importance of publication: the only way to spread the results of research throughout the world is to have them published.

Publication therefore has three objectives: to spread scientific findings, protect intellectual property and gain fame.

Scientists are obliged to publish their work, and publication justifies their existence. “A scholarly publication”, remarks Price, “is not a piece of information but an expression of the state of a

scholar or group of scholars at a particular time. We do not, contrary to superstition, publish a fact, a theory, or a finding, but some complex of these. A scientific paper is at the same time more and less than a concept or a datum or a hypothesis. If the paper is an expression of a person or several persons working at the research front, we can tell something about the relations among the people from the papers themselves" (Price, 1963).

Today, bibliometrics is applied to a wide variety of fields<sup>1</sup>:

- *the history of science*, where it elucidates the development of scientific disciplines by tracing the historical movements that are revealed in the results obtained by researchers;
- *the social sciences*, where, by examining scientific literature, it underpins analysis of the scientific community and its structure in a given society, as well as the motivations and networks of researchers;
- *documentation*, where it can count the number of journals per library and identify the journals that constitute the core, secondary sources and periphery of a discipline (by analysing the quantity of journals needed to cover 50 per cent, 80 per cent or 90 per cent of the information in a given area of science);
- *science policy*, where it provides indicators to measure productivity and scientific quality, thereby supplying a basis for evaluating and orienting R&D.

Bibliometric techniques have evolved over time and are continuing to do so: the counting of papers with attribution by country, by institution and by author; the counting of citations, to measure the impact of published work on the scientific community; the counting of co-citations (the number of times that two papers are cited together in a single paper); etc. All of these techniques combine to give more detailed and more effective measurements. Results are presented in various forms, such as mapping, in order to depict the relationships between participants and expand the means for analysis.

## CHAPTER 2. THE ADVENT OF BIBLIOMETRICS

### Background

The idea of examining literature goes back to the beginning of the century. In 1917, Cole and Eales published a statistical analysis of the history of comparative anatomy. The date was a milestone in the history of bibliometric analysis, as Cole and Eales were among the first to use published literature to build up a quantitative picture of progress in a field of research. Their paper describes the contributions of bibliometrics and the problems that it poses – some of which have yet to be solved.

Further work was carried out by Hulme (1923), this time using patents. By correlating patents and scientific literature in order to measure social progress in Britain, Hulme pioneered a modern methodology for the history of science.

Subsequently, Lotka (1926) showed the distribution frequencies of scientific production. He was undoubtedly one of the first to link the notion of productivity to counting, using the decennial indices of *Chemical Abstracts* and Auerbach's *Geschichtstafeln der Physik*. He also introduced a qualitative measure of scientific work based on data that made it possible to select the most eminent contributions. Lotka noted that the number of published papers was not distributed uniformly, and that productivity tended to be concentrated among a limited number of researchers.

In 1935, Cunningham published a study of biomedical literature and, in 1952, Boig and Howerton one of chemical literature. Until the 1960s, however, published research in this field was extremely rare. The fact that the term “statistical bibliography” was used fewer than five times between 1923 and 1962 illustrates how confidential such activity remained (Pritchard, 1969).

The 1970s brought a quantum leap in the number of bibliometric studies, crowning a second period in the history of bibliometrics which had begun with the advent of a database of citations of scientific papers, the *Science Citation Index* (SCI). Founded by Eugene Garfield in Philadelphia in 1963, the SCI paved the way for all those seeking to measure science using quantitative and objective methods.

Garfield's initial idea was to give researchers a quick and effective way to find published articles in their fields of research (Garfield, 1968). But he soon extended his work to evaluation of the references compiled: “It is concluded that as the scientific enterprise becomes larger and more complex, and its role in society more critical, it will become more difficult, expensive and necessary to evaluate and identify the largest contributors” (Garfield, 1979b). Garfield sought to portray citation analysis as a legitimate and practical tool for the evaluation of scientific production.

The SCI's existence not only sparked a large number of bibliometric studies, it also favoured the emergence of a new generation of bibliometrists claiming their discipline as the “Science of Science” (Price, 1965). Derek de Solla Price, an influential advocate of this methodology and a physicist by training, tried to take an approach to science that was independent of the one adopted by scientists. According to Price, science could be measured by publication, and it could be analysed independently of

scientists. Scientists, he reckoned, were specialists who, outside their respective fields of research, were no longer specialists. He wrote: "Just as economics has become a valuable aid to decision-making in government and industry as well as an academic subject in its own right, it may be that we are witnessing the birth of a similar scientific appraisal and analysis of the world of science" (Price, 1964). Price forecast that, in the near future, citation analysis would be used as a companion to peer review.

In this field, Russian researchers going back to the 1930s associated scientific analyses and the social sciences, for the purpose of providing methodological descriptions of the various disciplines. The systems of measurement they developed led to the establishment of a new field, *Naukometrika* (literally, "the measurement of science"), the forerunner of bibliometrics.

The international journal *Scientometrics*<sup>2</sup> was created with the aim of publishing papers on all the quantitative aspects of the "science of science"; it publishes a substantial share of contemporary bibliometric methods and studies and constitutes a highly active forum for discussion – at times quite intense – between representatives of the various schools of bibliometricians.

The first generation of bibliometricians created concepts and technical measurements that were later refined for use in the evaluation of science. But according to Wade (1975), these concepts had already been put to practical use prior to 1975, e.g. for evaluating the policies of "research councils", analysing university-level research and assessing needs for new research institutes in emerging fields.

## **The need to evaluate research**

Industrialised societies were highly favourable to the development of science (Bush, 1960). Since they began in 1957, the surveys conducted by the National Science Foundation (NSF) have demonstrated the American public's belief that science and technology make a major contribution to the progress of society (National Science Foundation, 1989, pp. 170-172)

Stimulated by competition with the Soviet Union, the United States made a considerable R&D effort in the 60s – one that involved the creation of various agencies and institutions. Similar moves then took place in Europe, the Soviet Union and Japan.

A change took place in the 1970s: science was no longer seen as a venture in which society could invest generously and without limit. The first phase of this shift was prompted by the slowdown of economic growth, but it also stemmed from a more critical attitude which took account of the negative consequences of scientific research: science and technology were expensive, but investment in research did not automatically make it possible to solve environmental problems or social problems such as the gap between the industrialised countries and the Third World.

This led to concern over the profitability of basic research in particular, and researchers were increasingly perceived as "producers of science" who had to account for the funds they received.

In addition, student revolt caused the image of universities to deteriorate, along with the authority of scientists and graduates. Such events aroused the suspicion of the general public towards science and technology. The new aim was to produce value added that conserved natural resources and created less pollution, and to create a more efficient research system that made better use of existing intelligence. It was in this context that the evaluation of scientific research came into its own.

As a result, the methods of the social sciences and the humanities (the "soft" sciences) were used to analyse "hard" sciences; quantitative criteria and measures were needed. In other words, methods had

to be found to quantify, compile and compare indicators. The establishment of a measurement of science(s) became inevitable. This shift cleared the way for the analysis of science and technology and favoured the advent of bibliometrists in science policy.

### **Bibliometrics and the measurement of science**

Governments in all countries have gradually perceived the need for critical analysis of their science and technology policies. Some have deemed it sufficient to create administrative units within their research ministries (Ministries of Education, Industry, etc.). Others have preferred to train specialists and develop indicators in an academic context which encourages the interplay of ideas.

In the United States, the National Science Foundation published its first *Science & Engineering Indicators* in 1972. In presenting their work, the officials in charge explained that “The ultimate goal of this effort is a set of indices which would reveal the strengths and weaknesses of US science and technology, in terms of the capacity and performance of the enterprise in contributing to national objectives. If such indicators can be developed over the coming years, they should assist in improving the allocation and management of resources for science and technology, and in guiding the Nation’s research and development along paths most rewarding for our society” (NSF, 1972). In subsequent reports, the role of bibliometric indicators expanded considerably.

Since then, bibliometrics has increasingly been oriented towards science policy. Groups of bibliometrists from different schools have proposed various methods for measuring the growth of science and linking their methodologies to evaluations. As a result, bibliometrics is now entering the difficult phase of trying to make a contribution to evaluation. Many of these methods have been presented at specialised seminars around the world.

In half a century, bibliometrics has thus earned its place as an instrument for measuring science, in Western industrialised countries as well as in Eastern Europe and industrialising countries such as India. In 1989, the OECD devoted a chapter of the Frascati Manual supplement to the higher education sector, confirming the place of bibliometrics in science analysis (OECD, 1989, pp. 49-53).

The Netherlands and the United Kingdom were among the first to publish regular studies on science using bibliometric indicators. Research groups in these countries were pioneers in theoretical construction and practical application in this field (Irvine and Martin, 1980; Martin and Irvine, 1984; Leven, 1982; Moed *et al.*, 1983), incorporating, in the 1980s, bibliometric measures into science policy analysis. In 1987, the Japanese Ministry of Education, Science and Culture (Monbusho) commissioned a team of bibliometrists to carry out a comparative study of the number of scientific articles published in seven major countries, in order to “set up an indicator permitting a better understanding of Japan’s research activities in the international context” (Ministry of Education, Science and Culture, 1987).

However, in all countries it has taken a long time for the bibliometric approach to gain acceptance as a measurement of science, in political as well as scientific circles. Some scientists continue to be hostile. Many of them have not acquainted themselves with the methodology and are uneasy with being analysed (or even evaluated) by a quantitative measure of the level of research activity – a “coarse-grained measure of the level of research activity, amounting to what some might say is merely ‘paper counting’” (Rappa, 1989, p. 28). It is not easy – psychologically or intellectually – to make the leap from measurement on the scale of a country or scientific discipline to the evaluation of individual researchers; it has even been called “an intolerable scandal” (Chauvin, 1991, p. 782). Some scientists have proposed a method of evaluation whereby job applicants would be judged not on a list of publications taken from a

somewhat anonymous database, but on select articles that they themselves deem most representative of their work. Researchers consider peer review the only way they should be judged. This view is shared by a large proportion of the scientific community at the present time (Ourisson, 1991).

At the level of individuals, bibliometrics measures the productivity of research but does not necessarily say anything about quality or the competence of researchers as teachers. Scientists' reactions are quite natural and underscore the need for interaction between the people being evaluated and the ones who are doing the evaluating.

In the evaluation process, the dialogue between "creators of science" and "bibliometrician-analysts" can be constructive. It can help alter the data and the methods used, but above all it can affect the interpretation of the results. Many experts in "evaluative bibliometrics" confirm that a discussion of the results must always be part of a researcher's evaluation (Moed *et al.*, 1983). In the dialogue between analyst and the person being analysed, checking the figures is essential. Clearly, the use of bibliometric indicators requires far greater vigilance for an individual evaluation than for a general description of science at the country level.

With time, these methods have become more widely recognised, but perhaps more so outside scientific circles than within the scientific community itself (Chelimsky, 1991). A large number of countries publish statistics along the lines of the NSF's *Science and Engineering Indicators: Australian Science and Technology at a Glance 1990* (Australia), *Science Indicators Compendium 1991* (Canada), *Science and Technology Indicators 1991* (Japan), *Science et Technologie - Indicateurs 1992* (France), *S&T Indicators Report 1994* (Netherlands), *Science and Technology Policy - Review and Outlook* (OECD), and the European report on science and technology indicators.

Apart from *Scientometrics*, the number of journals that publish articles using bibliometric methods is increasing. Among them, *inter alia*, are: *Research Policy*, *Science and Public Policy*, *Research Evaluation*, *Journal of the American Society for Information Science* and *Rapport de l'Observatoire des Sciences et des Techniques* (France).

## CHAPTER 3. BIBLIOMETRIC DATABASES

### The main bibliometric databases

The source for bibliometrics is always a database. Various bases, established by businesses or by public or private institutions, are used to illustrate the results of science and technology activity (with raw data). With special processing, they can be used to establish bibliometric indicators. Most databases are specialised; only a few are general in scope. Among the most widely used bases (see the Annex for greater detail) are:

- *Chemical Abstracts*: a specialist physics and chemistry database produced by an American company, Chemical Abstracts Services, for the American Chemical Society; it records an average of some 500 000 references a year taken from around 10 000 journals.
- *Compendex*: a specialist engineering and technology database produced by an American company, Engineering Information; it records an average of some 150 000 references a year taken from around 4 500 scientific journals.
- *Embase*: a specialist medical sciences database produced by a Dutch company, Excerpta Medica; it records an average of some 250 000 references a year taken from around 3 500 journals.
- *Inspec*: a specialist physical sciences database produced by the Institute of Electrical Engineers in the United Kingdom; it records an average of some 200 000 references a year taken from around 2 200 journals.
- *Pascal*: a general database covering several fields and produced by the Institute for Scientific and Technical Information (INIST) at France's National Centre for Scientific Research (CNRS); it records an average of some 450 000 references a year taken from around 6 000 journals.
- *Science Citation Index*: a multidisciplinary database produced by a US concern, the Institute for Information Science (see below).

The most frequently used sources of patent data are Derwent Information Limited's WPI(L) databases and that of Computer Horizon, Inc. (CHI). The Derwent databases are multidisciplinary and international in scope, recording patents issued and patent applications published by 30 national patent offices, whereas the CHI database draws mostly on statistics of the United States Patent Office<sup>3</sup>.

These databases are generally available on-line and/or on CD-ROM.

## **Problems posed by databases in general**

The choice of a database for compiling bibliometric indicators hinges directly on the objectives pursued and the questions the base must answer. Each database has its own content and entry criteria, and no two bases are identical. On any given subject, the quantity of articles (or other units of measure) will vary, depending on the database used. Quality (*e.g.* exact breakdowns by scientific discipline) will also differ, *inter alia* according to the journals from which data are drawn. For the same study, this diversity can yield divergent results, making it imperative to seek data that are as coherent as possible. When results differ according to the sources used, there is no objective means of distinguishing which of them most accurately depict(s) the reality of scientific output. All users of bibliometric indicators must therefore begin by choosing the databases best suited to their particular needs; to do so, they must first analyse the strengths, weaknesses and limitations of the various databases.

For “macro” level bibliometric studies, the bases selected must be representative, but they do not necessarily have to cover all of the data. By combining or factoring information from a variety of sources (databases) into the analysis, the risk of not being representative or exhaustive can always be minimised.

In addition, the bibliographical records that constitute the various databases were established to provide information of a primarily qualitative nature, and not as the basis for any sort of publication or article count. It is for this reason that the data, once extracted, need special processing to make them usable for the production of indicators. To this end, bibliometrists have developed a variety of processing methods and applied them in their analytical work.

In order to study the development of science in a particular discipline, it is necessary to propose or carry out an aggregation of the scientific field, because databases are not automatically classified by speciality. If the field to be analysed is “ceramics”, for example, analysts could query the database: *i*) by the titles of specialist journals in the field (identified by the analysts themselves); *ii*) by key-words connected with their research (also selected by the analysts themselves); or *iii*) by a select list of journals used by specialists. Depending on the query mode, the results vary. Attributing a search to a particular field or scientific discipline is a delicate task, especially insofar as a search program may very well bring together a variety of disciplines.

## **Types of literature**

Likewise, it is up to database users to decide what the count should include and exclude, depending on the study they are conducting. A typical database encompasses several different types of literature: articles, notes, summaries, letters to the editor, reports, notices, discussions, books, etc. For obvious reasons, articles are the basic mode of expression for conveying new knowledge. But for almost all the remaining types of literature, the choice lies with the analysts, who are required to choose the database contents that are to be incorporated into their bibliometric studies. On this subject, it has been noted – not without irony – that “one can easily fit the data with curves which show decline, increase or stability” of the science, depending on the types of literature chosen (Leydesdorff, 1991).

Users thus choose their counting methodology, data processing techniques and basic concepts. Consequently, the use of bibliometric indicators requires extreme prudence, and it is necessary to compare results obtained from several databases, especially if the scientific disciplines under study are fairly leading-edge or recent, and hence as yet relatively unstructured.

## **The Science Citation Index (SCI) database**

### ***Structure and potential of the SCI database***

In order to measure the quantity of a scientific stock, data of uniform quality must be used. A database must be built on defined and measurable criteria, so that analysts can specify the community being examined. It is in this context that the *Science Citation Index* (SCI) database, created by the Institute for Scientific Information in the United States, comes into its own.

Its “inventor”, Eugene Garfield, regarded cost-benefit considerations as paramount in defining the coverage of a database, conceding that it was not feasible to cover all existing journals. “One reason is that no one knows how many journals are published, because there is no agreement on what constitutes a journal” (Garfield, 1972).

Garfield first calculated the number of scientific journals needed to optimise coverage of a maximum amount of scientific information. To do this he adopted a law that had been developed by information scientists on the basis of Bradford’s work (1950). This law showed that between 500 and 1 000 journals were needed to cover 95 per cent of the “significant” literature published in a given field.

Garfield subsequently combined Bradford’s dispersion law with a concentration law he himself had developed (Garfield, 1972). Bradford’s law defines a scientific field, but if a database is to cover several such fields, does the number of journals for one field have to be multiplied by the number of fields? According to Garfield, because a substantial proportion of disciplines overlap, the core literature for all of these disciplines can also be covered by approximately 500 to 1 000 journals.

The first step was to develop a method for identifying these 500 to 1 000 journals. To do so, Garfield used the “number of citations” as one of the criteria for a “significant” search. Because authors cite earlier work in order to support, describe or develop a particular point in their own work, the citation of a scientific paper is an indication of the importance that the community attaches to the research. Thus, citations can be considered a criterion for selecting the most highly esteemed scientific journals on the basis of the articles they contain.

First, a count was made of the number of times an article was cited in a given journal. Then, the impact factor was computed by dividing the number of citations by the number of articles contained in the journal. This made it possible to eliminate any bias stemming from a journal’s size, rendering citation proportional to the number of articles.

Thus the SCI database covers the most widely used, recognised and influential scientific journals in the world, as measured by their “citation indices”. It limits the scope of coverage to world-class scientific journals, representing the “core” scientific output in specific fields and eliminating research not presented in the “mainstream”, which is limited to a specific group of journals. The problem for a database is to draw the boundary between “the strong” and “the weak”. Is it quality, quantity or other criteria that make the difference? In any event, for researchers, it is international acknowledgement of their output that determines the ranking of their worth. Around 1981, the SCI covered about 3 100 scientific journals; at the time, there were some 70 000 such publications worldwide! It was for that reason that the SCI environment was dubbed the “mainstream” (Frame, 1977). The selection of journals covered by the SCI thus injected a qualitative aspect into the literature count.

Apart from its virtual monopoly, the reason why the SCI is currently used so intensively cannot be attributed to the use of citations alone. The database covers a huge area of science; it is multidisciplinary, whereas most of the other bases are more highly specialised. With the SCI, it is possible to undertake a broad study of science, thanks to its uniform treatment of the exact and natural sciences; but ISI has also expanded its collection of databases, and it is now possible to obtain good information on social sciences and the humanities from the *Social Science Citation Index* (SSCI) and the *Arts & Humanities Citation Index* (AHCi).

The SCI also records the affiliation (addresses, institutional connections, etc.) of all authors of each article, whereas most other databases record only the first author of co-signed articles (the initial listing can sometimes constitute a “place of honour” for renowned researchers who have not necessarily contributed very much to the work). At the present time, *SCI* and *Physic Brief*<sup>4</sup> are among the rare databases to follow a policy of multiple listing. And yet this method offers a number of compelling advantages, especially at a time when studies of the internationalisation of science are far advanced. Because all of the affiliations of co-authors are recorded, a computer program can select an article written by researchers from different laboratories and different countries. This gives the SCI a standing in bibliometric research that bibliometrists cannot disregard.

### ***The Science Citation Index and its limitations***

#### ***Why do authors use citations?***

However, the SCI also has its limitations, which are common to many bibliometric databases. The reasons that prompt the author of a scientific article to cite other literature are complex. Science sociologists have been analysing the significance of citation for years and have pointed out that reference to the work of other researchers is not always related to the originality, importance or even the “quality” of that work.

Being cited may also depend on an article’s ability to reach a large audience. Famous scientists frequently supervise a large number of students, and their articles are more likely to be cited than those of their less influential counterparts (especially if the name of the “boss” appears – rightly or wrongly – among the co-authors). The weight of the social structure within the discipline is not necessarily directly linked to the quality of research. Authors may refer to eminent scientists as a tribute (or a tactic) rather than in acknowledgement of a piece of work they admire (*The Economist*, 18 January 1992, p. 87). The work of a researcher possessing an experimental technique or methodology, not necessarily of high quality but merely useful, will be cited each time it is used.

Citations are a measure of the overall impact of an article’s influence, or that of its authors, on the scientific community; they are a complex socio-epistemological parameter which probably induces a quality factor, but this factor is neither equivalent to, nor unequivocally correlated with, scientific quality (Seglen, 1992).

#### ***“Negative citations”***

References may also be negative. An author may be cited for research of a controversial nature or for an error of methodology. Here too, citation does not always measure the quality of research but rather the impact of a particular piece of work or of an individual scientist.

### *The “uncited”*

Furthermore, the number of scientists cited is extremely limited. Over half (55 per cent) of the articles published in the scientific journals covered by the SCI are not cited a single time in the five years following their publication. The uncited rate varies by discipline: in the engineering sciences category, the proportion exceeds 72 per cent (Pendlebury, 1991).

### *“Self-citation”*

The problem of self-citation, i.e. of researchers' references to their own past work, will be examined briefly later on.

### *The language factor*

Another point that has been frequently mentioned, criticised and on occasion analysed, is the fact that this database clearly favours English-speaking scientists (Otsu, 1983; Kobayashi, 1987). According to Garfield, it is neither easy nor cost-effective for the SCI data base to include journals that do not use the Roman alphabet (Garfield, 1975). This remark indicates that, in order to be accepted in the vanguard, articles have to be written in a mainstream language; clearly, the accent today is on English, and the system is self-perpetuating (Garfield, 1988). Moreover, researchers in non-English-speaking countries who publish in English enjoy a comparatively wider presence, as is the case in Scandinavia (Sivertsen, 1991).

### *The breakdown by scientific discipline*

One of the primary advantages of bibliometric databases is the possibility of using far more detailed and disaggregated classifications of scientific disciplines than the ones customarily employed to survey R&D expenditure and personnel. Normally, citations refer to articles, etc., within a single category or sub-category of a given field or discipline.

A problem emerges, however, when the SCI is used to study the state of science from a multidisciplinary database which covers a wide range and cannot reflect the various dynamics and specificities of citations from one discipline to another. *Theoretically*, selection criteria are the same for each of the journals covered in the database. Nevertheless, traditions and habits of publication and citation vary by field, and this affects the representativeness of data. For example, most (published and cited) world literature in physics, chemistry and biomedicine is well represented in the SCI. In contrast, there are problems with coverage of geosciences, biological field research, engineering and technology, mathematics and, to a certain extent, clinical medicine. The reasons for this are to be found in the fact that, in certain disciplines, communication is concentrated in the main specialised international journals, as well as in the fact that certain journals have a narrower “influence”. The language problem is also greater in certain fields of science.

All this applies to citation as well. To determine, say, a country's standing in terms of citations, it is necessary to factor in the specificities of each field, which affect the citations index. For example, it has emerged that, on average, in the short term, biomedical articles are cited more frequently than articles in mathematics or clinical medicine. Such correlations need to be taken into account, especially when figures are being interpreted. Corrections are needed, but they require a fuller understanding of the structure of the fields of science being analysed. This calls for a certain degree of prudence in interpreting the numbers; co-operation between practitioners of the discipline and bibliometricians would appear particularly necessary in such cases.

## **CHAPTER 4. USING BIBLIOMETRIC INDICATORS AND THE PRECAUTIONS TO TAKE**

### **Introduction**

Bibliometric analysis uses numerous parameters, such as scientific literature (articles, etc.), co-authorship, patents, citations, co-citations and co-words. These parameters are indirect measures of the scientific community, its structure and its output. Examples of how these indicators are used can be found in Chapter 5.

Bibliometric data and analysis provide information on the scientific orientation and dynamism of a country (or some other unit), and on its participation in science and technology worldwide – in other words, on its impact on both the national and the international community. Co-operation analysis makes it possible to identify and represent scientific networks and to highlight links between countries, institutions and researchers, as well as the impact of major programmes [CERN (European Laboratory for Particle Physics), WHO (World Health Organization), etc.]. Bibliometrics also highlights the structure of scientific disciplines and the links between them. Bibliometric data and indicators can serve as tools, or at least as an aid for describing and expressing questions that arise in the world of science.

As in other fields, it is important to note that the indicators obtained from bibliometric databases should be put in perspective. Indicators are based on a comparative approach: absolute values are not indicative *per se*, but take on their full significance only in comparison with those of other groups.

Analysis should also incorporate as large a volume of data as possible, so as to allow statistical compensation for any bias that might affect each small entity taken separately.

The limitations of the data used in bibliometrics stem mainly from the various means of communication scientists use to convey information to each other apart from the usual channel of scientific journals. *Inter alia*, oral communication between scientists is not captured in statistics, nor are internal reports between universities, laboratories or research groups and reports between countries working together through committees, programmes or laboratories. Also slipping through the net are important monographs and, to an even greater extent, electronic communication between researchers, which is developing rapidly. All of the communication that is covered by “traditional” bibliometric methods therefore consists of exchanges that have been “formalised”; informal communication is not incorporated and probably never will be.

The traditional approach is even more restrictive with regard to anything that involves industrial or defence-related research. There are great lags in communication between science (primarily academic) and industry, because of industry’s desire to protect its discoveries (prior to patent applications in particular) and the fact that its findings are generally published in an abridged form. Articles published by industrial laboratories deliberately give a limited view of the aims of research, which are generally to create products or processes subject to commercial competition.

Furthermore, a large proportion of defence-related research (which is often linked to industrial research) is never included in customary scientific communication, despite its technological importance and the fact that it tends to be at the leading edge of basic research.

### **The problems of co-authorship – “whole counting” or “fractional counting”?**

One of the ambiguities of the bibliometric method is the diversity of counting methods. For example, the classification of “co-authored” literature (*i.e.* articles written by more than one person) has long been a major subject of debate among bibliometrists (Martin, 1991; Braun *et al.*, 1991; Leydesdorff, 1991; Kealey, 1991). How can the participation of authors in scientific work be measured when the work is of a co-operative nature? Can individuals all be assigned “full credit” for their shares or, if an article is written with nine other authors, for example, should they each be assigned only a tenth of a credit? Does a country engaged in three-country collaboration forge a “whole link” or only a “third of a link”?

In practice, when an article is co-authored by researchers from different countries, bibliometrists basically have two ways of assigning credit to the countries concerned:

- Some assign full credit, *i.e.* count “1” for each co-author country (whole counting method),
- Others divide co-authorship by the number of countries of origin of the authors and assign a fractional credit to each country (fractional counting method). This method is based on “mathematical logic”: in order to obtain a final figure of 100 per cent, each country’s credit must be shared, and “1” must be divided by the number of co-author countries for each international co-publication.

Each counting method surely has its own logic, but the way credit is divided up must be totally mastered by bibliometrists and understood by the people who use their output.

For scientists and politicians who use science indicators, whole counting is far more comprehensible and easy to interpret. “A share of 10 per cent [of a country] means in this sense, that 10 out of every 100 papers in the world have at least one contributor from [this country]. It is hard to explain, however, the ‘meaning’ of a 10 per cent share in the fraction scale”, which may be the result of adding up ten or more papers (Braun *et al.*, 1991).

Even more importantly, fractional counting assigns a lesser value to international co-authorship than to authorship of a national article, when national performance is counted. The more international partners an article has, the less credit is assigned to each of the countries involved. Why should a country be credited more, in bibliometric statistics, for a paper by national authors than for a co-authored article during the course of international co-operation (Leydesdorff, 1991)? It is precisely for these reasons that some bibliometrists contend that fractional counting is an inferior procedure; especially when the volume of data is substantial, they maintain that “equal counting of all authors is in most cases the best solution” (Van Raan and Tijssen, 1990).

It is difficult to choose between the two methods. However, as long as both are in use, analyses of science and technology at country or laboratory level may vary, sometimes with contradictory results. The bias that these counts accumulate in favour or to the detriment of certain countries may make international comparisons awkward.

## The problem of database coverage

Other debates centre on a fundamental aspect of bibliometric measurement: how is it possible to measure scientific publication trends over a number of years using databases that evolve from one year to the next? In retrospect, this problem, which is tied to the development of bibliometrics, would appear to be the result of efforts to make bibliometric tools more representative. Some scientific journals disappear, others change their names or merge, and, above all, new journals emerge. The Institute for Scientific Information (ISI) monitors these changes and regularly updates the list of journals covered by the SCI, creating an annual turnover of approximately 7 per cent (Garfield, 1979a). In 1964, ISI had included around 610 journals in the SCI; by 1981, the number had risen to 3 600. There has been a parallel increase in the number of articles, from 100 000 in 1964 to 500 000 in 1981 in the SCI (non-fixed journal set) (Institute for Scientific Information, 1981).

However, this increase may pose a problem, for some, when attempts are made to track national scientific performance over time. For publication counting purposes, it may seem preferable for the number and make-up of journals in a base to be stable over the review period, so that the measure of the state of science in a given country would be comparable from one year to another. Growth in the number of a country's references could merely be the result of the addition of new journals to the base and not a measure of actual productivity growth. Data must be processed in a way that renders them comparable from one year to another (Anderson *et al.*, 1989). In order to make it easier to interpret the data, it was decided to track a constant number of journals among all those in the base – the “fixed journal set” representing approximately 2 100 of the publications monitored by ISI in 1973. Thus, the list of journals remained “stable” between 1973 and 1980. The list of selected scientific journals was again revised in 1981, to reflect new developments in science. New titles incorporated into the SCI were not added to the fixed journal set during the “freeze” periods (1973-1980 and 1981-1986); however, journals that were stricken from the SCI were taken out of the base.

Limiting references to a given number of journals would inevitably underrepresent the natural dynamics of knowledge, and this “freeze” could create an “artificial” world of science<sup>5</sup>. It may not be possible for new fields of research, such as supraconductivity or AIDS, which are discussed in journals that are created as a result of the relevant discoveries, to be studied immediately. The performance of industrialised countries is better represented in the non-fixed journal set because scientists in these countries tend to choose new titles, rather than traditional journals, to present their work. It is precisely in these new journals that new fields of research make their debut (Kealey, 1991). Thus, “the representation of research [provided by the fixed base] is significantly conservative and static” (Callon and Leydesdorff, 1987).

Here too, then, there are two different ways of processing SCI data in order to measure scientific performance: *i*) on the basis of a non-fixed journal set, with quantitative variations; and *ii*) on the basis of a fixed journal set, with the set diminishing over time. Only the starting points (1973 and 1981) are fixed; subsequent results do not evolve in the same way. Neither of the series is either “stable” or “constant”. The crucial question is to know which of the sub-sets is the more representative, or produces the more reliable indices for measuring national performance in science (Martin, 1991). In some specialities, the discrepancies are so great that trends vary over time (The Royal Society, 1986).

Counting papers is not difficult; making sense of the figures is more complex. The numbers do not “speak” for themselves; they need to be interpreted, taking into account the real and artificial bias in the data and in the method used to count them.

## **The marketing of bibliometric data and analysis**

Another difficulty encountered by bibliometrists is directly related to the commercial pressures that have emerged in the field.

There is a market for this type of study; bibliometric data and analysis can sell, and it is not uncommon for data extraction, processing and analysis contracts to involve substantial amounts of money. Each of the various bibliometric schools has developed its own method, and for this reason results can vary. Marketability has increased competition, but at the same time market pressure has encouraged professionalism in the field.

At the present time, bibliometrists are engaged in a lively debate over methodology. Methods are often similar, but there are neither uniform standards (Glanzel and Schoepflin, 1994) nor consensus as to the best methods or applications of bibliometrics.

## **What does the future hold for bibliometric indicators?**

Long restricted to evaluation and analysis of academic research and of major public programmes (both national and international), bibliometrics is gaining ground in other sectors, thanks in particular to the development of indicators to track various types of co-operation (Nederlands Observatorium van Wetenschap en Technologie, 1994; MERIT, 1994; and Katz and Hicks, 1996).

The basic indicators of bibliometrics still have a long way to go. Bibliometrists are pursuing their efforts to apply and improve existing indicators. One of the most interesting areas of study involves scientific and technological forecasting – a field for which indicators have been developed (Leydesdorff, 1995; Noyon and Van Raan, 1995). Work is also underway to develop “integrated indicators”, i.e. the association of several indicators to represent scientific and technological activities (Niwa and Romizawa, 1995). In addition to this macroscopic approach, “micro”-level trials are attempting to represent the shifts of science towards technology by looking at the development of networks of research groups (Hirasawa, 1995). Another focus of effort is to expand methods of evaluation aid based on the quality of scientific journals (Magri and Solari, 1996). These endeavours should be of particular use in analysing science policies.

Each indicator has its advantages and its limitations. Care is needed not to regard them as “absolute” indices; they are complementary. The various bibliometric procedures and methods need to be used in combination, despite the sometimes contradictory results, for as long as they offer useful information and comply with scientific and professional standards. Despite its limitations, bibliometrics provides an essentially objective quantitative measure of scientific output.

## **CHAPTER 5. THE MAIN BIBLIOMETRIC INDICATORS AND THEIR APPLICATIONS**

### **Introduction**

This chapter presents the best-known bibliometric indicators, with practical examples taken from international and national literature. The examples have been chosen more for their illustrative value than for the up-to-date nature of their contents. They are each accompanied by a brief commentary and methodological remarks.

The first part is devoted to the main quantitative indicators of science and technology activities, whereas the second deals mostly with so-called “relational” indicators, i.e. indicators that measure links and interactions among the various players in S&T systems, especially from the international perspective. There is also a brief presentation of techniques for “visualising” scientific variables via methods of multidimensional analysis.

### **Indicators of science and technology activity**

#### ***The number of papers (Examples 1-4)***

This indicator reflects scientific output, as measured by “paper” count – with “paper” used here to designate various media for scientific texts (books, journals, newspapers, reviews, reports, articles, etc.).

#### ***Uses***

Paper counts provide an initial, simplified and approximate measure of the quantity of work produced by a scientist, a laboratory, a school, a national and/or international R&D team, a country, etc. The number of such papers, in itself, constitutes a rough bibliometric indicator, but it is only by holding these basic data up against other “masses” that more significant measures of the *relative* impact of the subjects under study can be obtained. In this way, in a particular field or discipline, the research dynamic of a given country, team, etc. can be monitored and its trend tracked over time. Subsequent division of basic data by the number of researchers, or by amounts invested, yields “derived” indicators, which can to some extent enable the “productivity” of the work in question to be analysed.

#### ***Limitations***

It is reasonable to use the number of papers as an indicator when the numbers involved are high: the representation of a country, university, laboratory, field of research, etc. improves as this number increases. It would be less advisable to measure the output of an individual researcher solely by means of this type of indicator, inasmuch as such a measure does not factor in the quality of papers, even if the selectivity of the journals in the base is taken into account. Papers can also represent vastly different durations, volume and quality of research work. Such indicators cannot gauge the quantity or quality of the work represented by an article, and if the article is co-signed by a number of persons, the role of each one is known only to the authors themselves.

### ***The number of citations (Examples 5 and 6)***

Citations may be considered a measure of the impact of the articles cited, as well as of their timeliness and utility. It is presumed that a paper must have a certain quality in order to have an impact on the scientific community.

#### ***Uses***

Authors cite one another for a variety of reasons. Basically, citations may be divided into two groups: in one, earlier work is used to highlight the innovation contained in the article; in the other, the author acknowledges and pays homage to earlier work.

Citation data can also shed light on the interface between certain fields of science and technology (co-citations are discussed later).

#### ***Limitations***

Authors tend to cite work that is produced by their own scientific community, and that of authors who are “in vogue”. There are many reasons that prompt an author to choose between major works and to cite one rather than the other, and these reasons are impossible to identify. The works in question have not necessarily influenced their research.

Researchers can also cite their own work, thus increasing the number of citations with which they will be credited. “Self citation” is a very real phenomenon, and it lends credence to criticism of the reliability of such a measure. The problem is nonetheless rather minor if the volume of figures being analysed is high.

### ***The number of co-signers***

The number of a paper’s co-signers (or co-authors) is an indicator of co-operation at national or international level (internationalisation).

#### ***Uses***

Co-signature analysis is used to identify co-operation via papers that are signed by at least two different researchers. It can measure the volume of work carried out by teams of scientists at the individual or institutional level, as well as at the national or international level. It is a parameter for measuring the growth (or decline) of co-operative research as compared with research undertaken by a single scientist. Chronological co-signature analysis is one measure of the inroads of international co-operation into the “production” of national science and technology.

#### ***Limitations***

Theoretically, the affiliation (address) that is used is that of the researcher’s (or researchers’) place of work and not that of their residence or home country. The way in which a paper is identified and recorded depends on how that affiliation is listed – which can pose a problem. This is especially true in respect of institutional affiliations; some institutes and laboratories that are run by, say, very large universities or research organisations (such as CNRS in France) may be listed in databases under different names.

Another difficulty in tallying co-signatures stems from the diversity of counting methods (the “fractional” versus “whole” counting method issue examined in Chapter 4). Depending on the approach

taken, the results differ. The problem can be one of how to deal with a number of co-authors (from different countries) of a single article, or of how to handle a single author having a number of different affiliations, as in the case of a scientist temporarily seconded from his own laboratory to carry out research in a foreign institution.

In this latter case, if the author indicates the host laboratory only, that laboratory will get all of the credit for the paper. If both affiliations are listed, however, the paper will be credited to both institutions (and countries) and will take on the appearance of an international co-signature. But such complications are not very significant in the aggregate.

### ***The number of patents (Examples 10 and 11)***

Patent statistics provide elements for measuring the results of resources invested in research and development activities, and most particularly trends in technical change over time. Patents constitute an initial form of legal protection for the inventions developed by firms, institutions or individuals and, as such, may be considered an indicator of inventiveness. Patent statistics are increasingly being used as science and technology indicators, and patent documents contain a number of elements that can be used in bibliometric analysis.

In 1994, the OECD released a manual entitled *The Measurement of Scientific and Technological Activities. Using Patent Data as Science and Technology Indicators. Patent Manual 1994* [OCDE/GD(94)114]. The Organisation's S&T databases also contain a number of series on patents.

OECD statistics (see *Patent Manual 1994* and *Main Science and Technology Indicators*) cover patents applied for (as opposed to those granted or issued) under national, European and international procedures.

Four types of patent data are involved:

- the number of “resident applications” from inventors living in the country concerned over the course of a given period. This indicator gives an idea of technology output; from it can be drawn additional information, such as the “coefficient of inventiveness” (i.e. resident applications per 10 000 inhabitants).
- the number of “non-resident” patent applications, submitted by inventors not living in the country concerned. This indicator reflects technological penetration.
- the number of national patent applications, which is the sum of resident and non-resident applications. In a sense, this indicates the size of the “technology market” that the country represents.
- the number of patents applied for abroad by inventors residing in the country concerned. This indicator reflects the country’s “technology diffusion”.

### ***Uses***

Patent counts can be used to situate an invention and the role of each inventor in the development of new techniques; they are therefore a measure of innovation and technological capacity at the level of nations, industries or firms. Initial work on patent statistics as S&T indicators focused on

clearly identified objects, such as molecules. Subsequently, competing technologies were measured, as was the level of inventiveness of countries in competition over a major invention.

US patents were first used as indicators of output. Patent applications in the United States are subjected to a thorough analysis of the originality of the claimed invention. One advantage of these documents is that they contain highly detailed descriptions and references for inventions and provide important information (such as citations) that is of bibliometric significance.

#### *Limitations*

The tendency of industrial inventors to seek patents varies by industry and from one firm to another; some major technological improvements do not lead to patents. Similarly, the “quality” of patents is not necessarily of the same level; not all patents have the same significance in terms of technical innovation and economic promise. It is therefore ill advised to compare patent applications for diverse technologies or different industries. Nonetheless, in a clearly defined macroscopic domain, such as that of countries, comparisons can be made. Despite their limitations, patents are and will be increasingly useful as a source of information to provide an approximate measure of innovation.

#### ***The number of patent citations (Example 12)***

This indicator measures the impact of technology (more, perhaps, than the impact of science).

#### *Uses*

There is still no commonly accepted method for measuring patents in terms of absolute or relative value, but patent citations can be used as an S&T indicator. The first page of a patent generally contains a reference to patents that have already been approved on the same subject. The patent inspector proposes these references during the examination process.

Citing patents is one way to depict the state of a given art, i.e. how what has already been done in similar fields relates to the newness and significance of a proposed invention. Because such patents are likely to be significant ones, citations may serve as an indicator of the importance of the cited patent to the technology for which protection is being sought. Applicants themselves sometimes furnish patent citations as part of their applications, but these are less commonly used in bibliometric analysis.

#### *Limitations*

The citations chosen by examiners raise questions about the reasons that lead them to cite references that differ from those cited by the applicants themselves. Examiners are not specialists, and they may cite patents more for their legal importance than for their innovative nature. Moreover, the citations proposed by patent applicants are not yet accepted as a truly significant measure of the importance of the patents cited, since the choice may have been motivated by factors other than scientific importance. The limitations of such measures need to be understood, since patents can be written in such a way as to conceal major inventions behind minor advances, in order to mislead the competition. Business enterprises, guided by their legal counsels, exhibit considerable diversity in the way they protect their research work.

## **Relational indicators (Examples 13 to 16)**

### ***Co-publications***

This indicator measures interactions and scientific relationships between networks, teams, institutions and countries.

#### ***Uses***

A co-publication is the result of co-operation between representatives of each entity and each country taking part in a particular joint research programme. Such research forges links between the parties (scientists, laboratories, institutions, countries, etc.) that have worked together to produce a scientific paper. The total number of links instituted by particular participants can be defined, depicted and measured by co-authorship. Using an indicator of co-authorship, it is therefore possible to outline these relationships. Following this principle, it is possible to construct a matrix with, in each cell, the number of co-signatures between the author (or authors) listed on the rows and the author (or authors) listed on the columns. This indicator can identify the main partners in research endeavours and provide a description of scientific networks.

#### ***Limitations***

The usefulness of these indicators is directly linked to how issues of affiliation and co-authorship counting are handled (problems dealt with in Chapter 4).

### ***The affinity index (Example 17)***

The indicator used to evaluate the relative rate of scientific exchanges between one country (A) and another (B), over a given period of time (and, if desired, in a specific area of science), and in relation to all international co-operation between these same two countries over the same period, is called the “affinity index”.

This indicator provides a “dual vision” of these links, which can be measured, for example, in terms of co-authored articles; clearly it can be applied to entities other than countries (e.g. businesses, geographical aggregates, etc.).

The formula for computing the affinity index, which was developed by the *Laboratoire d'évaluation et de prospective internationales* (LEPI) of France's CNRS, is as follows:

$$\frac{COP (A - B)}{COP (A - WD)} \times 100$$

where COP (A-B) represents the number of scientific links (co-operation) between A and B and COP (A-WD) is the number of co-operative links between A and the world.

#### ***Uses***

This indicator measures not only the links between countries, but also their equilibrium level, i.e. the “balance of power” underlying the flows; it shows the strongest areas as well as the weakest. An examination of how affinity indices vary over time yields an indication of changes in bilateral scientific relations.

### *Limitations*

Affinity indices cannot be applied unless there is a certain “mass” of co-operative links during the period – links running in both directions. It is preferable to use this indicator in respect of scientific co-operation between two parties having similar scientific mass.

### ***Scientific links measured by citations (Examples 18 and 19)***

This indicator measures networks of influence between scientific communities.

#### *Uses*

Citations can be used to trace networks of influence between different scientific communities. Such interactions highlight peer evaluations of past and ongoing scientific work.

#### *Limitations*

A number of problems with this approach were discussed in Chapter 4.

### ***Correlations between scientific papers and patents (Examples 20-22)***

This indicator illustrates links (interactions) between science (as measured by papers) and technologies (as reflected in patents).

#### *Uses*

Much information may be taken from patents and their accompanying documentation, such as references to scientific articles, some of which are included in specialist databases. The link between scientific knowledge (articles) and technologies that use that knowledge can be analysed by references or citations by inventors and/or patent examiners.

Two types of indicators have been put forward. The first links science and technology through scientific citations and patent citations. The second measures the length of time between the publication of scientific articles and patent applications.

The indicator of the intensity (or “scientific proximity”) of an industrial or technological activity is based on the relative number of citations of scientific articles in the patents applied for in that sector.

#### *Limitations*

Correlations between patents and articles have not yet been analysed systematically, and opinion is divided on their importance and significance. Patents primarily serve a legal purpose, and the fact that their authors seek both to demonstrate their technological links and conceal the essentials of their content undermines the credibility of any utilisation of such data for analytical and statistical purposes.

### ***Co-citations (Example 23)***

Co-citations measure the number of times that two papers are cited simultaneously in the same article. This indicator illustrates thematic networks and the influence and impact of authors. In the final analysis, the co-citations method represents the scientific community’s reactions to the results of research.

### *Uses*

Clusters of co-citations provide a description of similar and related research subjects and of complementary research in the speciality concerned, which is itself measured by citations. It is also possible to identify and “map” communities of researchers within particular networks. Such clusters also make it possible to show how fields and sub-fields evolve over time.

### *Limitations*

Because they describe only a part of the process of assembling knowledge, co-citations provide a highly selective analysis of science – one that refers far more to scientific literature than to technological literature.

### ***The “co-occurrence” of words (Example 24)***

The previous indicator (co-citations) looks at the number of times two articles are cited together. This indicator examines the frequency with which two given words (“co-words”) in a particular S&T field are used together in papers or patents. For each word, its co-occurrence with another word is analysed, along with its frequency. The words in question are specific to each topic of research and are selected by experts in the field.

The assumption underlying the method is that co-words can be used to identify and depict specific networks of a given type of research, with a view to studying their development. In scientific papers and patents, the presence of these words reflects a likeness of intellectual concepts among researchers. They are therefore like signals, indicating associations which can be represented in the form of lexical graphics (“leximaps”). The frequency of word associations is used to construct maps (strategic diagrams) that represent the major themes of the field under study, and relationships among them.

### *Uses*

This method has been used, for example, to describe the role of a government agency and to consolidate and transform a network in macromolecular chemistry.

### *Limitations*

The method raises problems of interpretation of the results. Words cannot be separated from their syntactical context, and there does not seem to be any systematic way of interpreting maps. Users stress the importance of micro-level analysis.

### ***Visual representation techniques for scientific fields and countries (Examples 25, 25A and 25B)***

Since it is difficult to capture and represent the structure of tables composed of many figures, a variety of methods based on techniques of multidimensional analysis (e.g. minimum spanning trees, correspondence factorial analysis, etc.) are used to construct maps that allow for various interpretations of bibliometric data for different purposes.

### *Uses*

These maps of relational networks make it possible to depict the structure of research in the various fields and sub-fields of science and to observe, with greater clarity than in statistical tables, all of

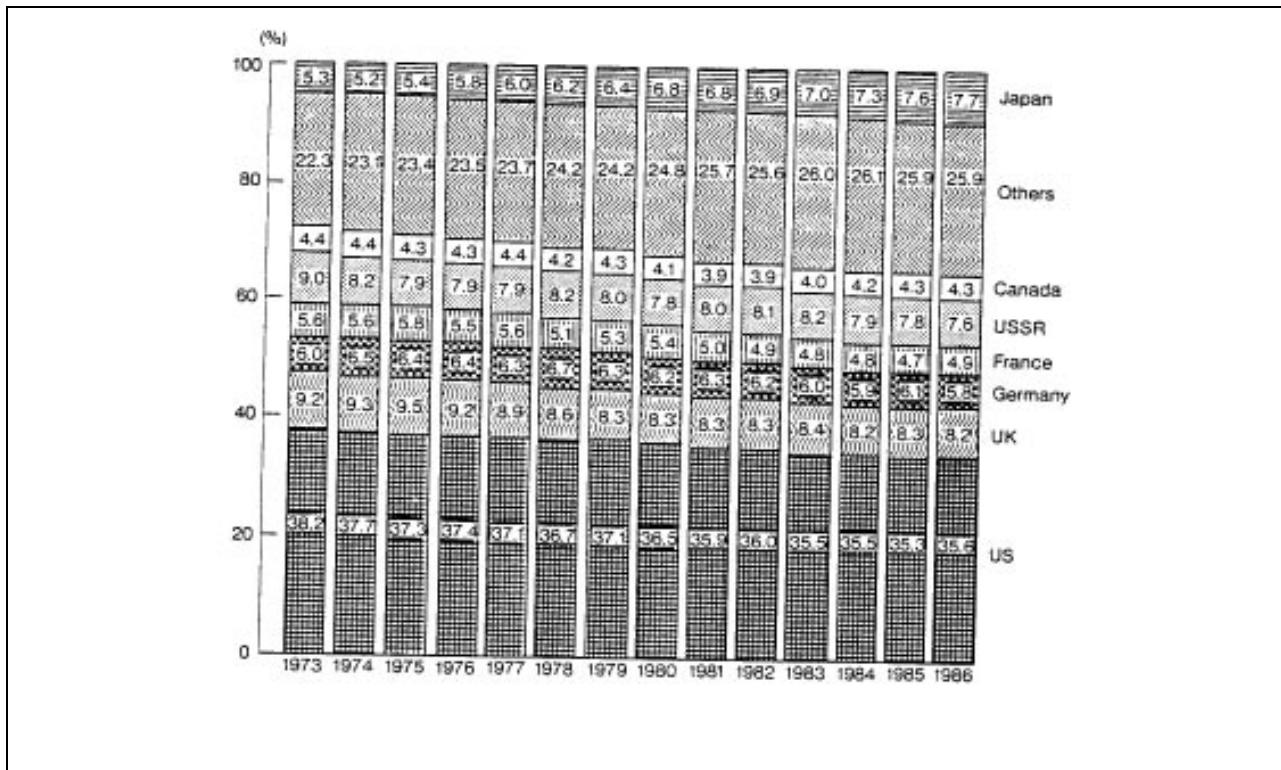
the links that have been forged between countries and/or fields of science. Thanks to these techniques, it is possible to situate the relative positions of different-sized countries in global scientific co-operation.

#### *Limitations*

Because it is not possible to represent all of the data contained in a multidimensional system in two dimensions without losing information, losses must be minimised by combining various techniques.

## LIST OF EXAMPLES

**Example No. 1.**  
**World scientific output 1973-86 by main country (%)**

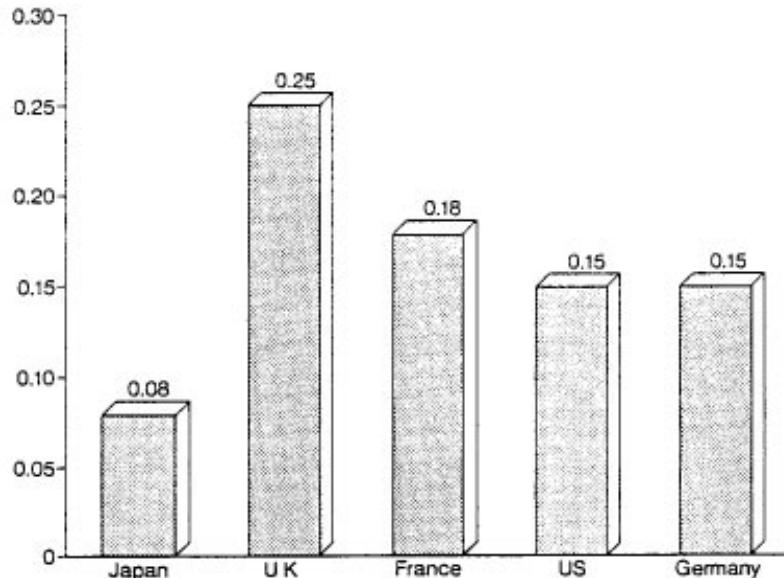


Source: Science and Technology Agency (Japan) (1991), p.13.

The breakdown of world scientific and technical literature by main “producer” country is one of the “classic” bibliometric indicators, as reflected, for example, in various reports by the National Science Foundation (NSF) in the United States, and in the Scientometrics series, which regularly publishes information broken down by country and by major field of science.

The above figure indicates that, in 1986, US articles accounted for 35.6 per cent of the worldwide total in “hard sciences” and “life sciences” (i.e. all except “soft sciences” such as social sciences and the humanities). Of the other large countries, the United Kingdom, Japan and the USSR each accounted for about 8 per cent of the total. Except for Japan and the group of “other countries”, which saw their shares increase over the period, a relative decline was observed for all countries.

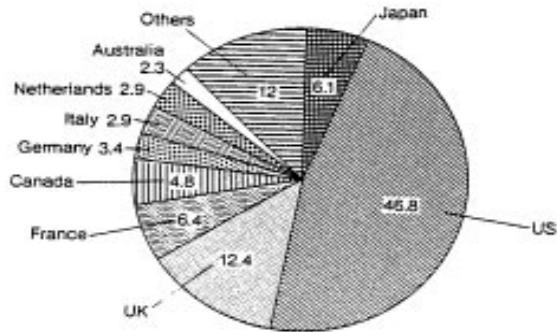
**Example No. 2.**  
**Number of papers per researcher in five countries, 1986**



Source: Science and Technology Agency (Japan) (1991), p.13.

This chart would suggest that “productivity”, as measured by the number of papers per researcher, was considerably higher in the United Kingdom and France than in Japan, the Japanese figure being only about one-third of the British average (0.08 in Japan versus 0.25 in the UK). To put these data into perspective, a number of factors that affect their comparability have to be taken into account. It must be noted that OECD statistics on researchers, which were used to compute these percentages, are in fact overestimated for Japan (whose data are closer to the total number of employees, whereas for other countries they are expressed in full-time equivalents). Furthermore, in Japan, a majority of researchers work in private industry, whereas the proportion of academic researchers is higher in most of the other countries (publication practices being fairly different between the industrial and higher learning sectors). Lastly, because the analysis is based on the number of papers in English, Japanese references are underrepresented.

**Example No. 3.**  
**Papers on the “human genome” by author-country, 1991 (%)**

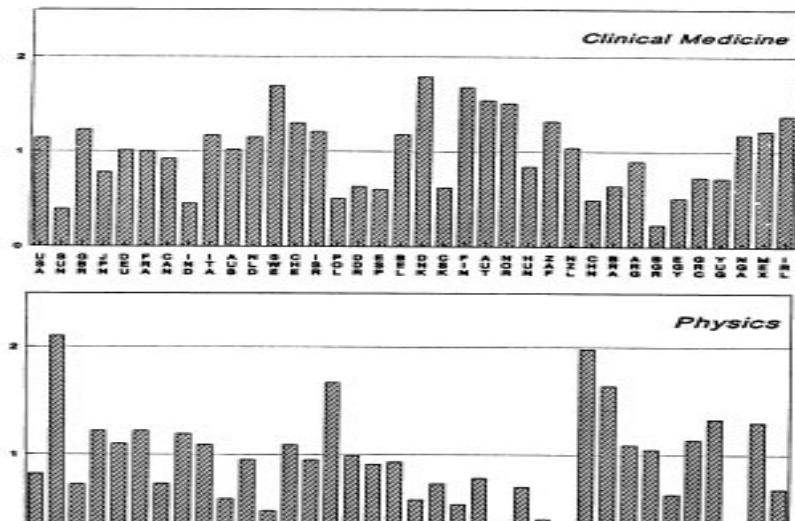


Source: Science and Technology Agency (Japan) (1991), p.14.

In life sciences, data on the human genome are considered a key to the understanding of organic functions. This figure presents the distribution by author-country of recorded literature on this subject in 1991 (using a slightly different breakdown from that of Example No. 1). A comparison of national contributions shows that the United States and the United Kingdom are the main participants, with shares well in excess of their respective contributions to scientific output in general (see Example No. 1).

### **Example No. 4.**

## **Specialisation by discipline: shares of clinical medicine and physics in national scientific literature, 1981-86**



*Source:* Miquel and Okubo (1994), pp. 271-297.

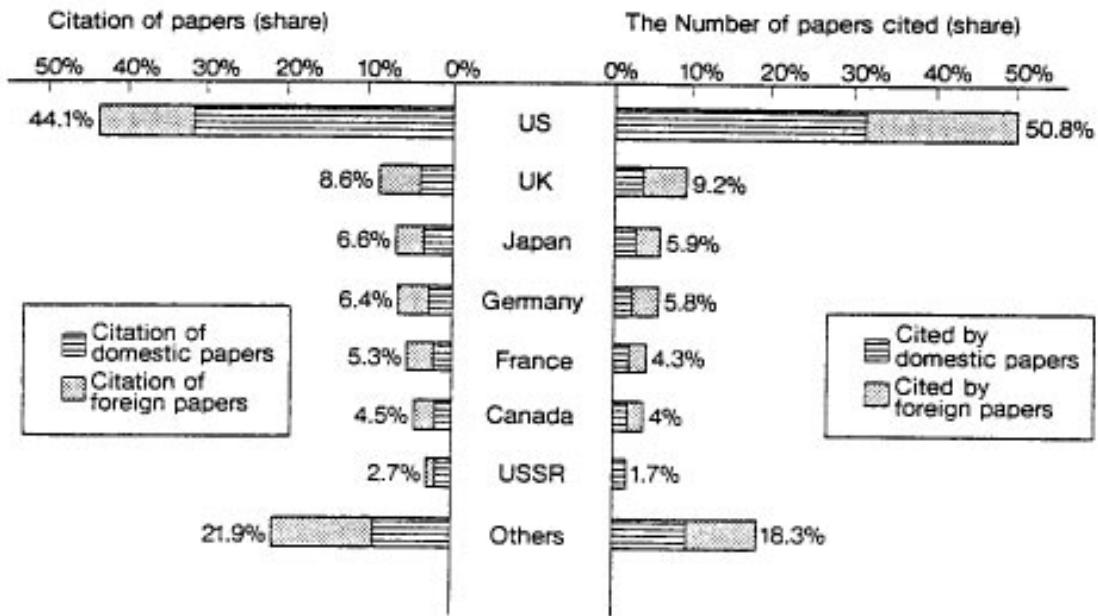
See Annex for abbreviations.

National production (as measured by the number of scientific papers) in the two fields of clinical medicine and physics, as compared with those same fields' share of all recorded literature (the "world reference"), is presented for the 36 leading "producer" countries.

Level 1 represents the “world reference” which is (or is not) reached by the countries; to a certain extent, this illustrates their “degree of specialisation”. Countries that approach level 2 (where the rate is double the “world reference”) therefore have highly pronounced specialisations.

Thus, there are sharp contrasts between countries as well as between the various fields of science (“strong” and “weak” fields), compared to the world profile of scientific literature. This is particularly noticeable in clinical medicine and physics. In clinical medicine, almost half of the countries have a higher proportional share (typically, Scandinavian countries and countries having cultural ties with the United Kingdom) than the “world reference”. In a sense, the histogram for physics is the inverse of the one for clinical medicine.

**Example No. 5.**  
**Percentage of papers (¹) citing and cited by other papers (¹),**  
**by main country (1984-86 averages)**



(¹) National and foreign papers.

Source: Science and Technology Agency (Japan) (1991), p.16.

Citations of national and foreign papers (world reference) are presented for a number of countries. It is generally held that foreign literature has the greatest significance (impact).

The left-hand portion of the diagram shows, for example, that about 44 per cent of all the citations in the world are made by US scientists; the bulk (about two-thirds) of those US citations concern work by other Americans, whereas researchers in other countries have a greater tendency to cite foreign literature.

The right-hand portion shows that papers by US researchers are the most commonly cited in the world (about 51 per cent) and that, as above, a great deal of the citations are by other members of the American scientific community. British scientists account for some 9 per cent of the literature cited.

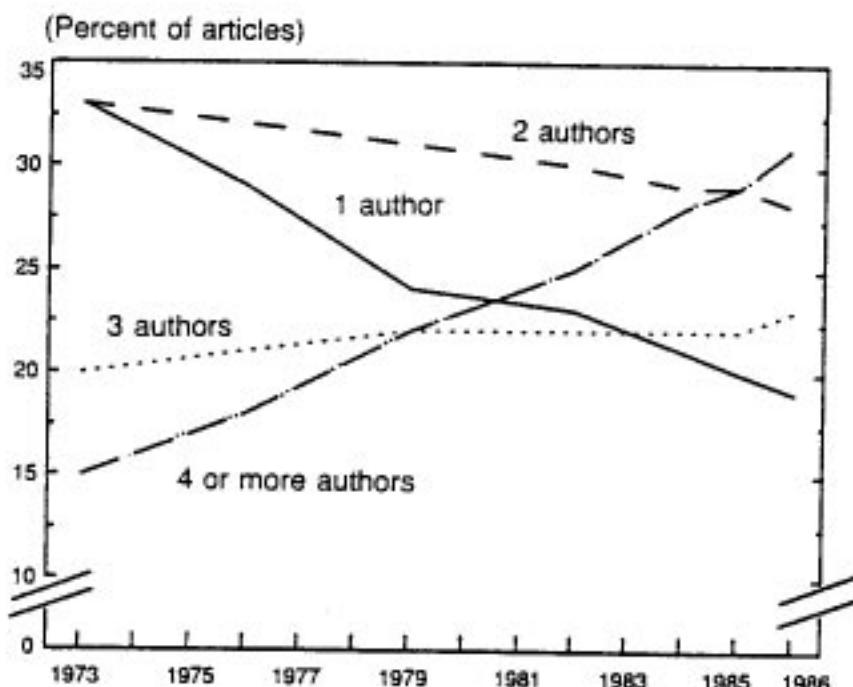
**Example No. 6.**  
**Published citations in 1984, by main field of science and by R&D orientation in engineering and technology**

<b>Category of research</b>					
	Total	Applied technology	Technological engineering and science	Applied research and targeted basic research	Basic research
<b>Number of citations</b>					
<i>Field cited</i>					
All fields	88 504	15 835	43 527	18 468	10 674
Engineering	59 483	15 093	40 599	3 660	139
Physics	14 501	441	647	8 787	4 626
Chemistry	7 605	62	452	4 443	2 648
Others	6 915	239	1 829	1 578	3 261
<b>Percentage of citations</b>					
All fields	100.0	17.9	49.2	20.9	12.1
Engineering	67.2	25.4	68.3	6.2	0.2
Physics	16.4	3.0	4.5	60.6	31.9
Chemistry	8.6	0.8	5.9	58.4	34.8
Others	7.8	3.5	26.4	22.8	47.2
<b>Percentage of world papers in the field</b>					
<i>Field</i>					
Engineering	100.0	41.5	50.6	7.9	0.0
Physics	100.0	1.0	2.4	32.5	64.1
Chemistry	100.0	0.7	2.8	27.8	68.7
Articles are put into the category of the journal in which they are published; the content of these journals ranges from "predominantly applied research" to "predominantly basic research". Except for the "Total" column, where percentages are read vertically, columns show percentages horizontally, i.e. they indicate the percentage of articles in each category of research, by field. The worldwide distribution of papers was determined for the period 1973-79. Data for later years suggest a similar pattern. Citations were counted for 1984 and earlier years (to 1973). Data represent articles, notes and reports in over 2 100 of the most influential journals contained in the <i>Science Citation Index Corporate Tapes</i> of the Institute for Scientific Information.					
<i>Source:</i> National Science Board (1989), p. 346.					

Citations are used here as indicators of the influence of basic research in the realm of engineering and technology ("hard sciences"). The table presents (in actual quantities and percentages) aggregate world citations since 1973. The papers in question have been categorised according to their orientation, which ranges from "predominantly applied research" to "predominantly basic scientific research" (columns) and by main subgroup of engineering and technology (rows). It can be seen that, in engineering and technology, papers tend to cite their own sub-area of science. In 1984, approximately 68 per cent of citations in engineering were attributed to other papers in the same field, whereas the proportions were much lower (under 6 per cent) for physics and chemistry. Journals oriented towards basic research in engineering and technology contained almost no citations of literature in the engineering subgroup (less than 1 per cent of the total), but far more in physics and chemistry (about a third of the total for each) and, above all (almost half, 47 per cent), other fields.

**Example No. 7.  
Science and engineering articles**

**having one, two, three, four or more authors, 1973-86 (%)**



Source: National Science Board (1989), p.120.

Between 1973 and 1986, worldwide, the average number of authors per article increased from 2.3 to 3.0. At the same time, the proportion of articles having a single author declined, from 33 per cent of the total in 1973 to 19 per cent in 1986. Over the same period, articles co-signed by four (or more) authors rose from 15 to 31 per cent of the total. This trend toward greater co-authorship can only reflect expanded co-operation between researchers and research groups, i.e. the development of larger research teams and research tied in with major national and international programmes ("big science"). However, a closer look at the data (see the NSF's publication statistics) reveals fairly major variations from one discipline to another. For instance, in clinical medicine and biomedical research, co-operation rates were the highest (with, in 1986, respectively 43 and 34 per cent of articles having four or more co-authors and only 14 per cent having a single author). In contrast, in mathematics, almost two-thirds (62 per cent) of the papers were the work of a single author, 30 per cent were written by two authors and 7 per cent by three. Only 1 per cent of all mathematics papers resulted from co-operation among four authors or more. The chart shows world scientific co-operation trends since the early 1970s in all fields of science combined.

**Example No. 8.**  
**Articles co-authored by academic and industrial researchers in the United States, in**  
**relation to aggregate articles authored by industrial researchers, by field of science**  
**1981-91 (%)**

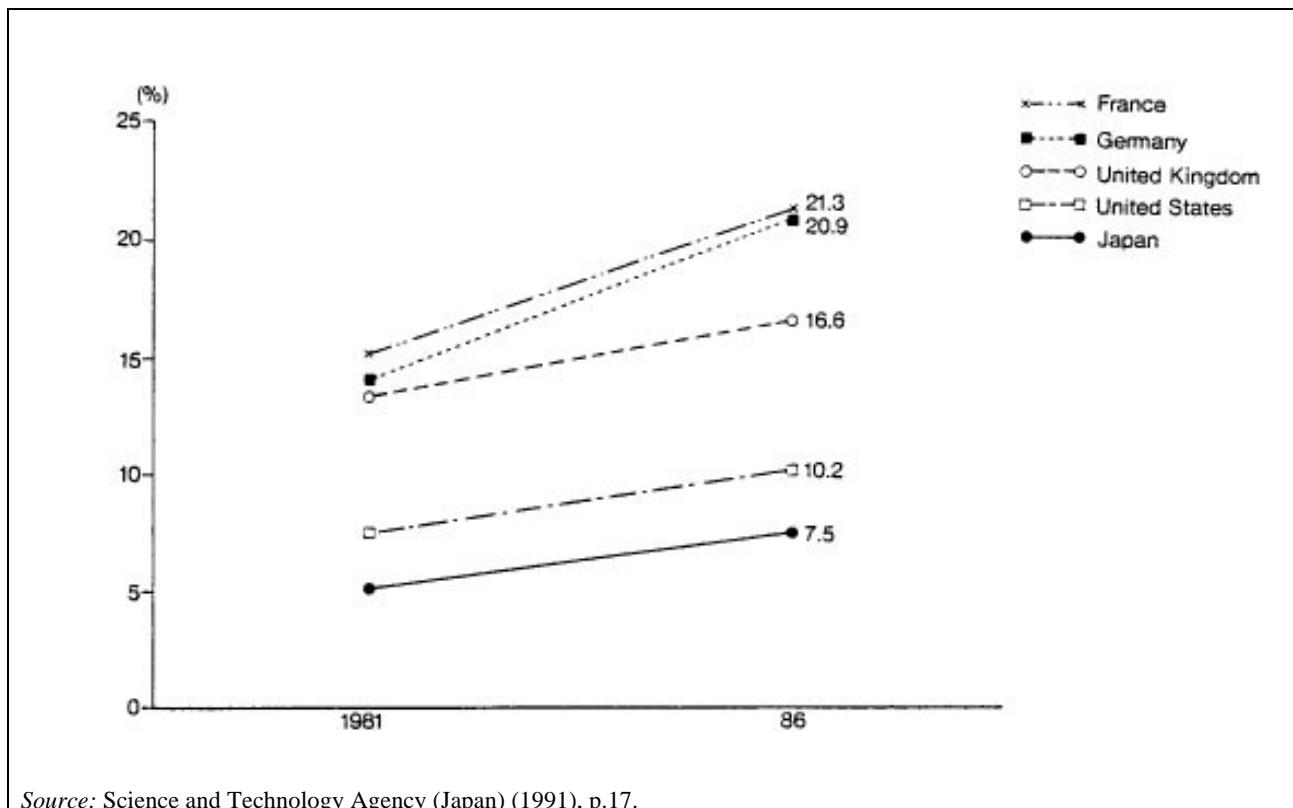
	198 1	198 2	198 3	198 4	198 5	198 6	198 7	198 8	198 9	199 0	199 1
<b>All fields</b>											
Clinical medicine	22	24	23	25	27	28	30	31	32	33	35
Biomedical research	30	34	33	35	40	37	42	41	42	44	45
Biology	35	37	35	35	39	38	40	41	39	39	40
Chemistry	39	46	42	37	44	44	41	47	48	43	45
Physics	13	17	15	16	16	18	20	20	22	22	24
Earth and space sciences	20	21	23	25	23	23	25	26	28	29	31
Engineering and technology	34	35	33	36	33	36	34	41	38	40	37
Mathematics	16	17	16	17	18	20	23	24	23	26	26
	43	35	42	42	43	40	42	41	51	52	49

Source: National Science Board (1989), p.429.

This table traces co-operation between academic and industrial researchers in the United States over the years from 1981 through 1991, as measured in terms of co-authored articles, by major field of science.

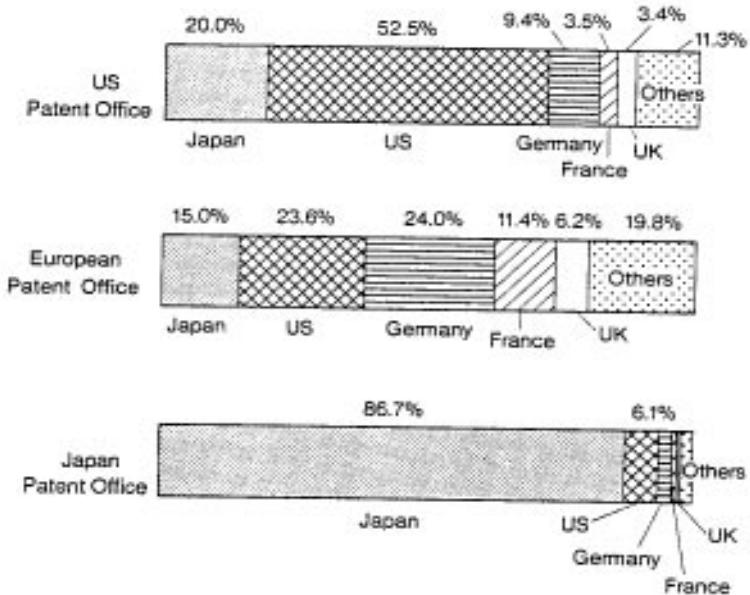
Such co-operation has risen sharply. In 1991, 35 per cent of all the papers authored by industrial researchers were co-signed by academics (versus 22 per cent in 1981). This trend has affected all disciplines, to varying degrees. Mathematics has the highest percentage of co-operation (nearly half of all the literature), along with life sciences (clinical medicine, biology and biomedical research), with rates of around 40 to 45 per cent. Literature in the fields of chemistry and engineering-technology shows the lowest ratios (respectively 24 and 26 per cent).

**Example No. 9.**  
**International co-authorship trends for selected countries, 1981-86 (%)**



The above data show the growing internationalisation of research. Between 1981 and 1986, the major countries expanded the proportion of their national literature having foreign co-authors. In 1986, about 10 per cent of all the scientific literature in the United States was co-signed by authors with foreign affiliations. At the same time, the proportions for Western European countries ranged between 15 and 20 per cent. Japan had the lowest rate of internationalisation, as measured in this manner.

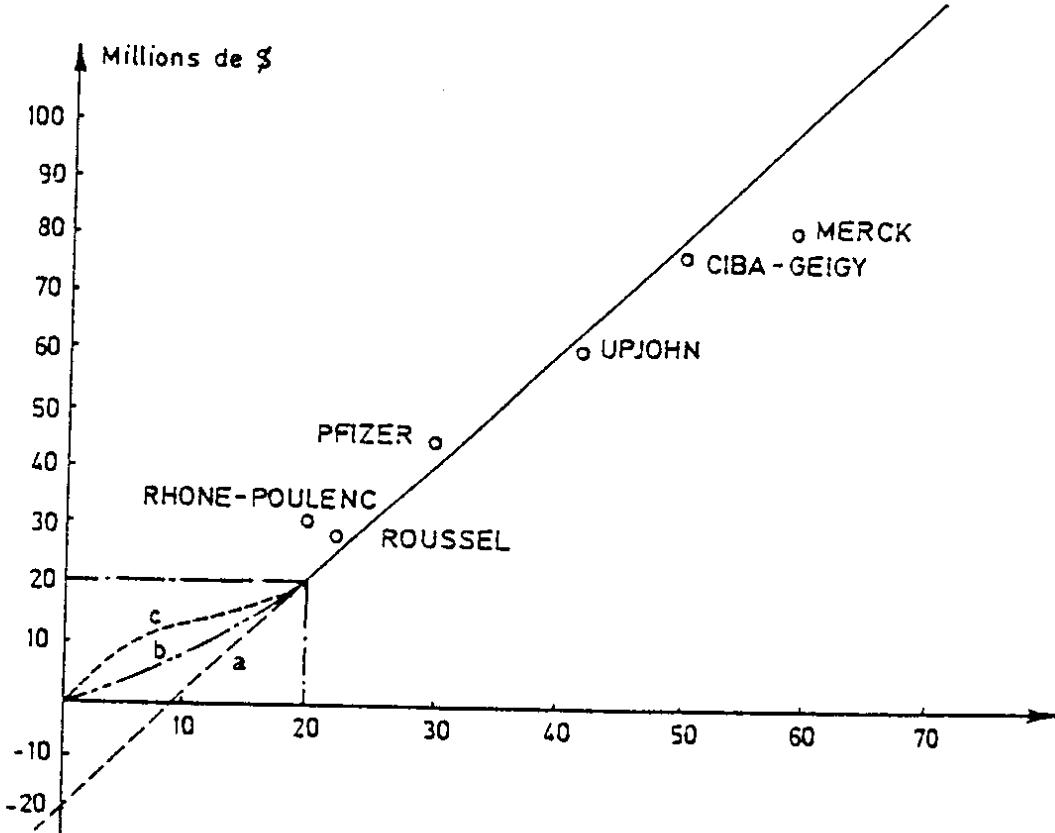
**Example No. 10.**  
**Patent applications to the Patent Offices of the United States, Europe and Japan,**  
**by main country (%)**



Source: Science and Technology Agency (Japan) (1991), p.13.

In the United States, residents accounted for roughly half (52.5 per cent) of all patent applications in the late 1980s. The rate was considerably higher in Japan, where applications by residents represented a very large majority (nearly 87 per cent) and those by non-residents about 13 per cent of the total. One fifth of the applications recorded by the US Patent Office and 15 per cent of applications to the European Patent Office came from Japan. US inventors appear far more inclined to seek patents in Europe than in Japan. The main European countries apply for far fewer patents in Japan than they do in the United States.

**Example No. 11.**  
**“Productivity” of industrial pharmaceutical research**



Source: Miquel and Doré (1981).

This diagram shows the “productivity” of the scientific activity of selected major pharmaceutical multinationals, in terms of the ratio of turnover (vertical axis) to the number of patents obtained (horizontal axis) in the 1970s; here, patents are considered a rough measure of the “results” of research.

The resultant regression line shows that, above a certain threshold, there is a relatively satisfactory correlation ( $r = 0.814$ ) between the two variables, demonstrating that patenting policy conforms to a “law” that is closely linked to research budgets. Different firms all follow this “law” in a similar manner, since they have very little latitude in managing their research in a highly competitive market in which scientific output is the key, not only to future earnings, but to survival itself.

**Example No. 12.**  
**Citations of US patents by subsequent patents, classified by inventor country**  
**and by sector holding the cited patent**

Year of patent cited	<b>a. Inventor country</b>																			
	United States	Japan	Nether- lands	United Kingdom	West Germany	Canada	Franc e	Switzer- land	Italy	Sweden										
Citations per citable patent																				
1975	3.91	4.03	3.63	3.63	3.33	3.33	3.09	2.98	2.99	3.24										
1980	3.09	3.35	2.90	2.70	2.51	2.50	2.41	2.37	2.34	2.28										
1985	1.37	1.84	1.15	1.21	1.15	1.11	1.07	1.10	1.04	1.02										
<b>b. Sector of the holder for US inventors</b>																				
All inventors			Firms		Government		Individuals		Foreigners											
1975	3.91		4.15		3.00		3.34		4.13											
1980	3.09		3.37		2.41		2.46		3.42											
1985	1.37		1.49		1.07		1.02		1.44											
Figures – especially those for recent years – will rise as patents continue to be cited.																				
Source: National Science Board (1989), p. 140.																				

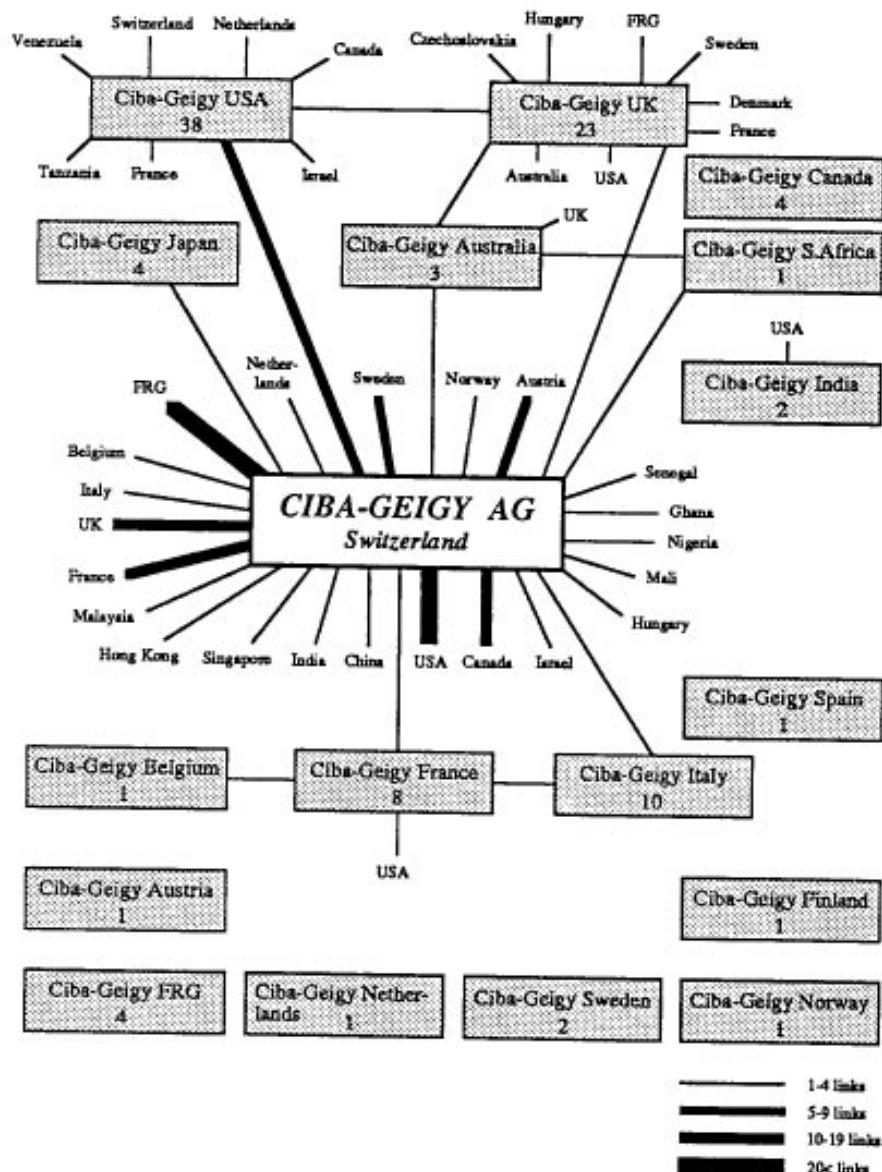
The above data refer to citations, by patents applied for in the United States in a given year, of other patents. They show the countries (and also, in the detailed tables of the NSF, the products and industrial sectors) from which patents receive the most or the fewest citations.

It should be noted, however, that the “structure” of these citations can vary considerably, not only between industrial sectors but also over time, because of “time lags”. In other words, the older a patent is, the more time it will have had to be cited, and the younger it is, the less its importance will have attracted the attention of other patent applicants.

Of the ten countries studied, Japan is the one whose patents were cited most over the period in question, followed by US patents and those of the Netherlands, the United Kingdom and Germany. This would seem consistent with the actual number of patents applied for by these same countries in the United States.

With regard to the sectoral distribution of patent owners, it is US firms that hold the most frequently cited patents. Patents belonging to the Government of the United States, or to individual Americans, are cited less. In 1975 and 1980, US firms had as many patent citations as holders of Japanese patents. Almost all of the Japanese patents obtained in the United States belong to businesses.

**Example No. 13.**  
**The international network of the Ciba-Geigy Corporation, 1990**

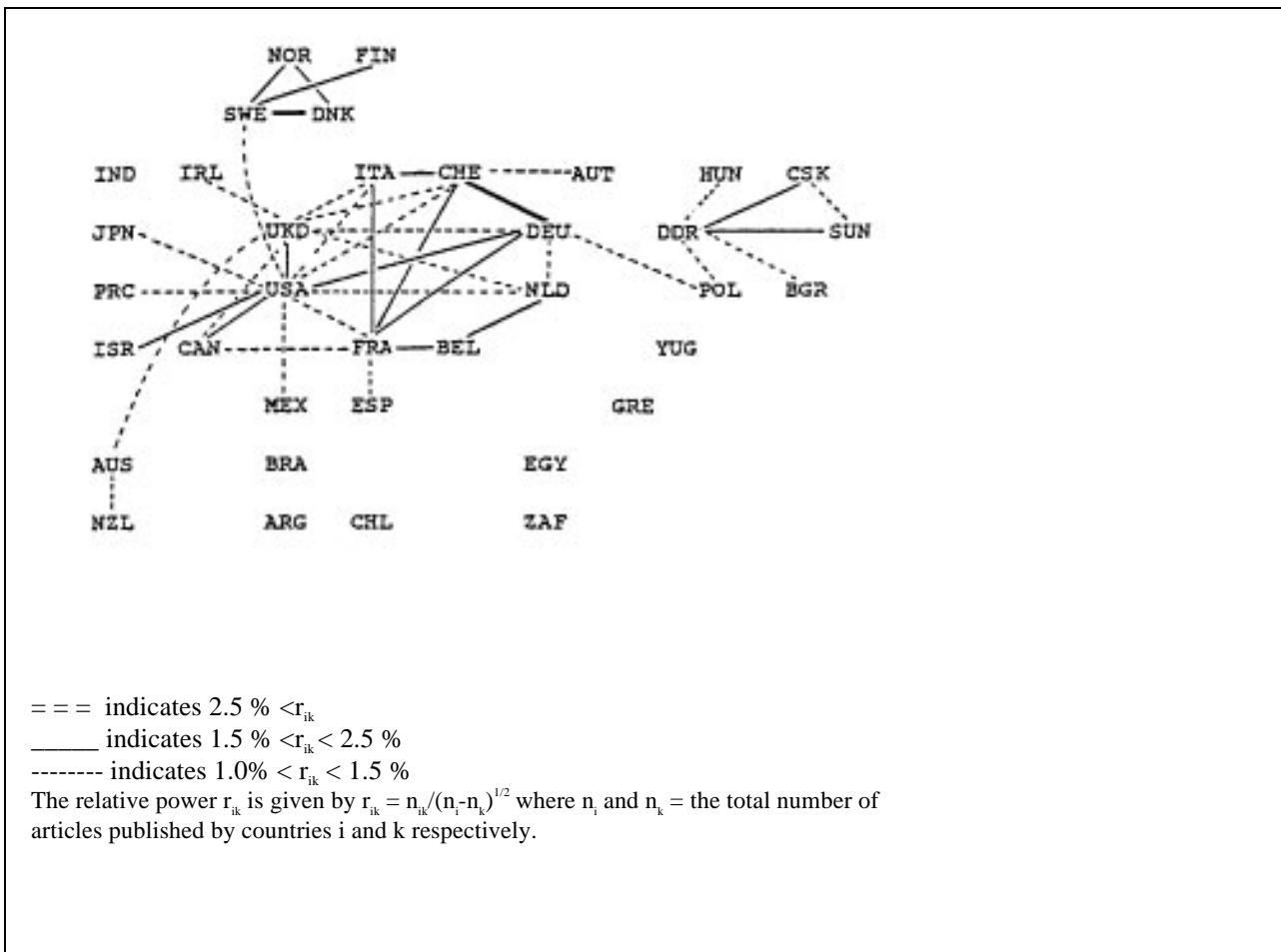


Source: Okubo (1993).

The grey rectangles designate subsidiaries that took part in scientific activity, and the figures show the number of published articles. Ciba-Geigy, a pharmaceutical firm, displays a high degree of internationalisation, since some 28 per cent of the group's scientific articles involved international co-authorship, and 32 countries were directly tied to Ciba-Geigy's centres of excellence, through the parent company in Switzerland or one of its subsidiaries. The degree of internationalisation of research, in relation to the group's total scientific output, was higher for the parent company (36 per cent of whose articles were co-authored). Nevertheless, subsidiaries in the United States and the United Kingdom also

contributed to this internationalisation by forging their own international ties with foreign universities and hospitals.

**Example No. 14.  
Map of international co-operative links in the sciences, 1981-85**



Source: Schubert and Braun (1990).

This diagram traces the main networks of international scientific co-operation for the 36 most "productive" of the 163 countries in the world that produced at least one internationally co-authored article during the period. It was constructed using Table 1 of the source article, which indicates, for all these countries, the total number of published articles, as well as the number of those articles having international co-authors; from this, ratios were calculated and expressed as percentages which show the degree of internationalisation. For these 36 countries, the authors observe four major clusters of varying size; the largest covers the United States and Canada as well as most Western European countries, the second the Nordic countries, the third the chief Eastern European countries and the fourth, which is very small, Australia and New Zealand. Linkages between the Eastern European group and the Western scientific community seem to involve Poland, and those of the Nordic group, with the United States in particular, involve Sweden. Australia and New Zealand have strong ties with the United Kingdom, and

Israel with the United States. Many of these configurations appear to reflect historical and/or political connections.

**Example No. 15.**  
**Japan's position in the international scientific activity**  
**of Germany, France and the United Kingdom**  
**in engineering (ENT) and earth and space sciences (EAS), 1981-86**

<i>Germany</i>		<i>France</i>		<i>United Kingdom</i>	
ENT	EAS	ENT	EAS	ENT	EAS
UNITED STATES	UNITED STATES	UNITED STATES	UNITED STATES	UNITED STATES	UNITED STATES
SWITZERLAND	UNITED KINGDOM	GERMANY	GERMANY	CANADA	AUSTRALIA
FRANCE	FRANCE	CANADA	UNITED KINGDOM	GERMANY	GERMANY
UNITED KINGDOM	NETHERLANDS	UNITED KINGDOM	CANADA	AUSTRALIA	CANADA
JAPAN	ITALY	ITALY	ITALY	FRANCE	FRANCE
	CANADA	POLAND	NETHERLANDS	CHINA	NETHERLANDS
	SWEDEN	BELGIUM	USSR	EGYPT	ITALY
	JAPAN	SWITZERLAND	SWITZERLAND	NETHERLANDS	SOUTH AFRICA
		EGYPT	SPAIN	JAPAN	NORWAY
		SPAIN	BELGIUM		INDIA
		JAPAN	SWEDEN		SWEDEN
			JAPAN		GREECE
					SPAIN
					SWITZERLAND
					JAPAN
<i>Japan's %</i>		2.6	2.0	1.8	2.5
Source: Okubo (1994).					

This table situates Japan's role, in terms of the number of scientific links established, in the overall international activity of the three leading European countries, in two large areas of science in which Japanese power is uneven. Germany, France and the United Kingdom each have their specificity in terms of international activity, which leads to contrasting strategies – either explicit or implicit – for the edification or acquisition of knowledge. The result is that, in Germany, Japan's importance is clearly greater in engineering than in earth and space sciences. Since Japan comes out higher in the field in which it is strong, there is a certain “rationality” to co-operation between Germany and Japan.

To some extent, this same correlation is to be found in co-operation with the United Kingdom; here too, Japan ranks higher in engineering than in earth and space sciences. Nonetheless, in engineering, the United Kingdom engages in a far greater volume of co-operation with Canada, Australia, China and Egypt than it does with Japan. In this case, historical experience would seem to play a more important role in the establishment of scientific links. France has not adopted the same “rational” logic as Germany and Japan. With regard to Japan, there is no significant difference between engineering and earth and space sciences; France works first and foremost with its neighbours.

**Example No. 16.**  
**Comparative profiles – Strasbourg and France, 1986**  
**Scientific co-operation links between Strasbourg and 97 partner countries**

Country	MAT	PHY	CHM	ENT	EAS	BIO	BIM	CLI	Total
United States		16	12	3	3	3	34	4	75
Germany		26	2	1	2	2	23	7	63
United Kingdom	1	9	8		1	4	17	2	42
Italy	1	11	10			2	8	6	38
Canada		1	3			1	6		11
Switzerland		17			2	2	14	5	40
Belgium	1	7	3			1	4		16
Spain	1	10	10			1	4		26
Netherlands		3	2				5		10
Sweden		6	1				5	2	14
...		...	...	...	...	...	...	...	...
Other					1		3	1	5
Number of links	4	158	57	11	20	21	142	28	441
Profile of Strasbourg (a)	0.9	35.8	12.9	2.5	4.5	4.8	32.2	6.4	100%
Profile of France (b)	2.7	30.7	11.8	3.2	8.3	4.3	18.0	21.0	100%
Comp. profiles (a/b)	0.3	1.2	1.1	0.8	0.5	1.1	1.8	0.3	

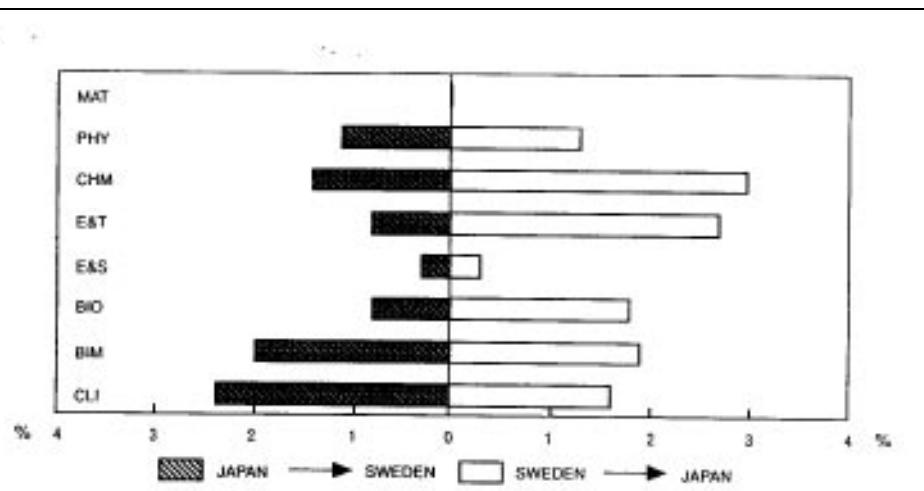
MAT (mathematics), PHY (physics), CHM (chemistry), ENT (engineering-technology), EAS (earth-space), BIO (basic and applied biology), BIM (biomedicine), CLI (clinical medicine)

Source: Okubo *et al.* (1992).

The city of Strasbourg is one of France's leading scientific centres. A comparison of Strasbourg's links with those of France as a whole shows the special features of the activities of this university town. Strasbourg has a proportionately high share of international co-operative links in physics: 158 links (or 35.8 per cent of the city's total links). In France as a whole, physics accounts for 30.7 per cent of total links. When the position of physics in Strasbourg's international ties is compared with the figures for France (the "comparative profile"), it can be seen that Strasbourg's numbers are somewhat (20 per cent) higher. The participation of Strasbourg institutions in CERN projects is one reason for this, since CERN accounts for 23 per cent of Strasbourg's co-authored articles in physics. This nuclear physics activity also affects engineering and technology, via methods and construction of dedicated instrumentation.

An exceptional proportion of international activity involves biomedical research (142 links, or 32.2 per cent of the total) and constitutes another distinctive feature of Strasbourg's scientific involvement, this ratio being above the national average of 18 per cent. The level of joint endeavour is due to the presence of a number of specialised laboratories (molecular biology-genetic engineering, cellular molecular biology, neurobiology, etc.), as well as to that of two industrial institutions that take part in international biomedical co-operation. In contrast, clinical medicine's share of Strasbourg's links (6.4 per cent) is particularly low, in relation both to the city's overall international activities and to clinical medicine in France as a whole.

**Example No. 17.**  
**Affinity indices between Japan and Sweden 1981-86 (%)**

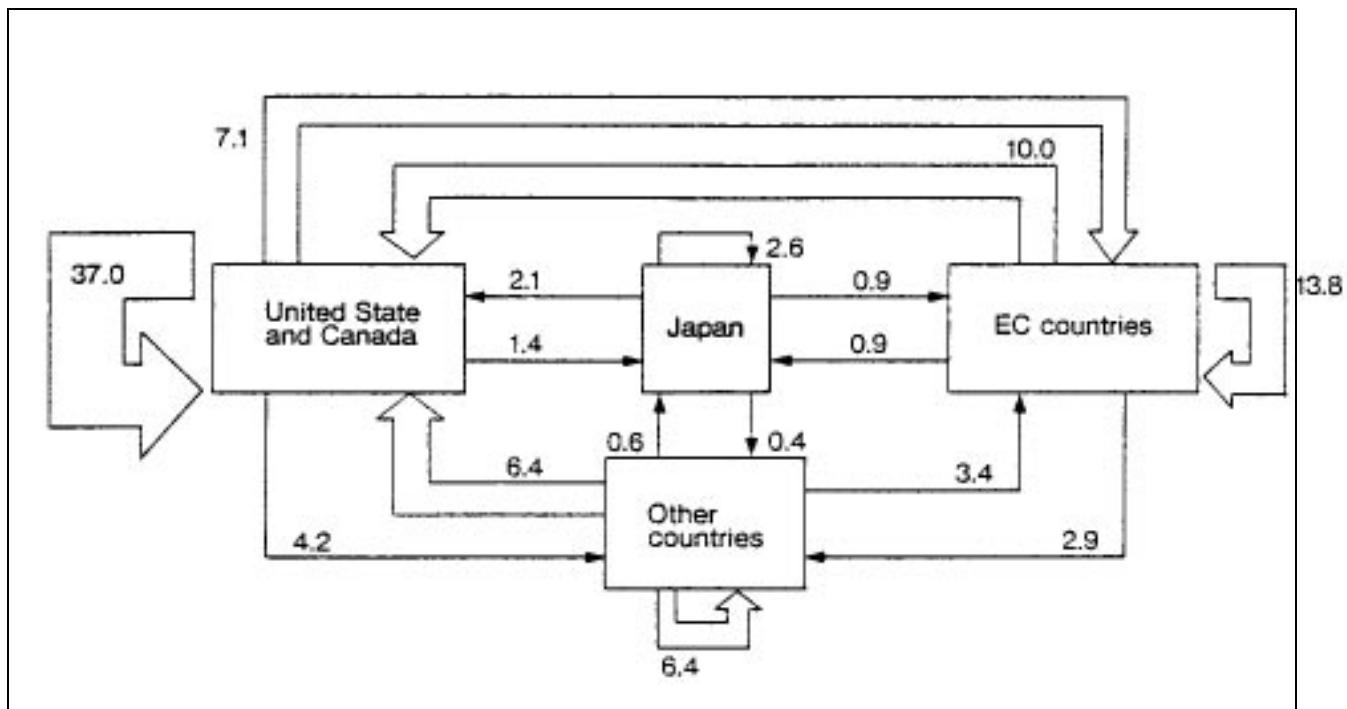


Source: Miquel and Okubo (1990).

This chart shows how affinity indices can be used to illustrate scientific co-operation between Japan and Sweden.

The co-operative links would suggest that each country favours the fields of science in which the other is stronger. Japanese co-operation with Sweden is particularly extensive in clinical medicine and medical biology (the predominant areas of research in Sweden), whereas Swedish collaboration with Japan is at its greatest in chemistry and engineering and technology – areas in which the Japanese are relatively strong. In contrast, there is hardly any co-operation in mathematics or earth and space sciences, where neither country possesses any exceptional abilities.

**Example No. 18.**  
**Worldwide distribution of citations, 1984-88 (%)**



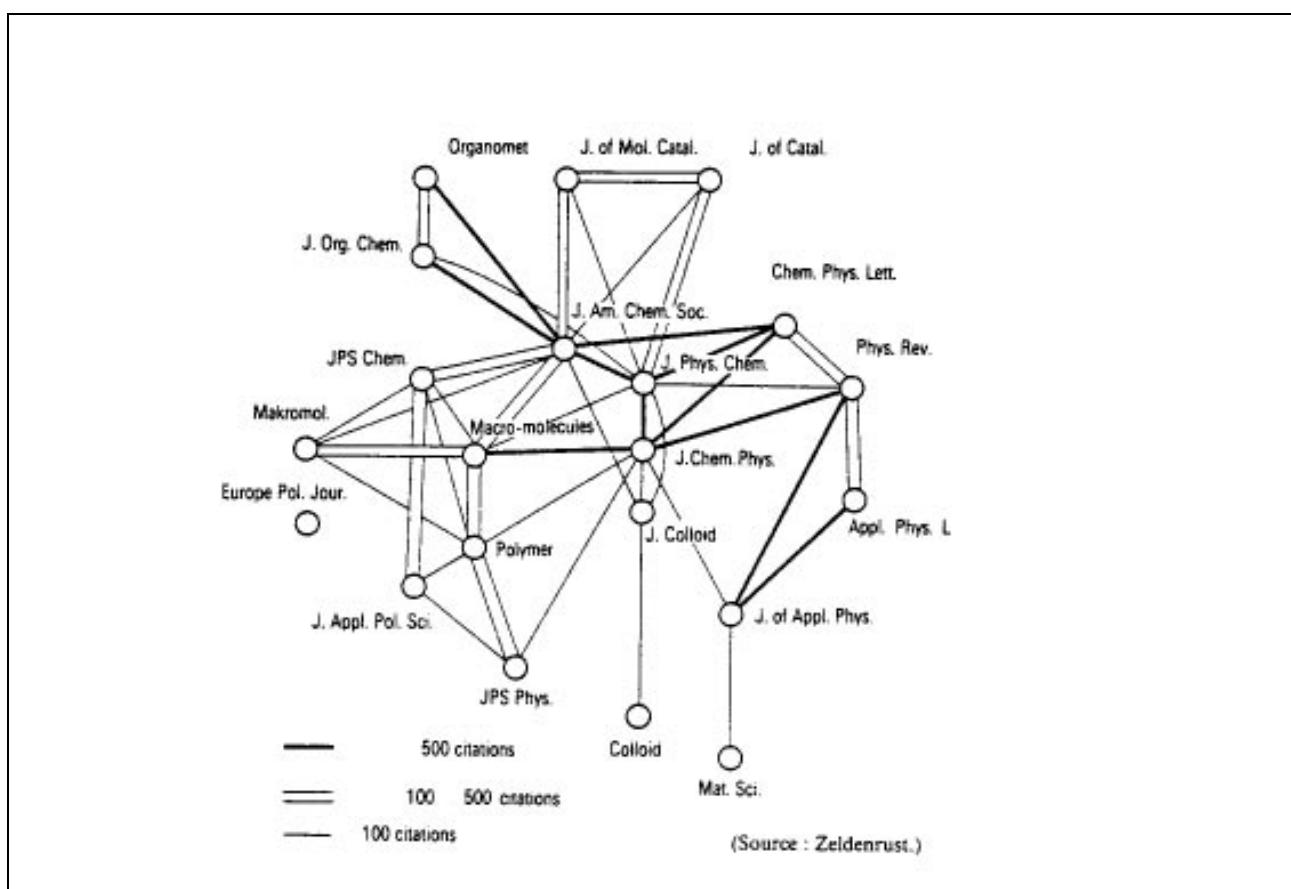
Source: Science and Technology Agency (Japan) (1991), p. 17.

This diagram illustrates the distribution of aggregate worldwide citations (in all fields of science combined) for 1984-88, by main (groups of) citation countries. It attempts to answer the questions: “Who cites whom, who is cited by whom, and is there any (im)balance between countries with respect to these citations?”

It shows, for example, that the United States and Canada “received” some 55.5 per cent of world citations, of which 37 per cent from articles published in those same two countries, 10 per cent from European Community countries, 2.1 per cent from Japan and 6.4 per cent from other countries. Similarly, the United States and Canada produced nearly half (49.7 per cent) of world citations; apart from self-citations, 7.1 per cent of these were of papers from EC countries, 1.4 per cent of Japanese papers and 4.2 per cent of papers of other countries.

It should be noted that the SCI database can be used to identify bilateral scientific links, with a more detailed breakdown, by disaggregated sub-discipline.

**Example No. 19.**  
**Citations between specialised journals in the field of polymers**



Source: Callon *et al.* (1993), p. 62.

This diagram outlines the linkages, as measured in terms of citations, between the main journals specialising in the field of polymers.

“The thickness of the lines suggests the intensity of the linkages, i.e. the number of cross-citations between articles appearing in these journals. The closer the linkages between publications, the more complementary their content and the more the clusters they form suggest the existence of a coherent and integrated discipline. Dynamic analysis of the clusters can be used to track shifts and reorganisations of these areas of research. It is necessary to allow for the fact that journals can see the contents of their articles evolve over time, and that the stability of clusters can mask significant thematic changes” (as translated from Callon *et al.*, 1993, p. 63).

**Example No. 20.**  
**Relationship to science of world patents issued in the United States**

Activity	Scientific intensity of patents		Age of articles cited (years)
	(Number)	(Average = 100)	
Electronics	0.31	29	7.7
Pharmaceuticals	3.19	298	7.8
Instruments	0.54	51	8.1
Machinery	0.12	11	8.2
Transport*	0.01	1	8.3
Chemicals	0.77	72	9.7
Other	0.12	11	8.5
Arithmetic average	1.07	100	-
All	0.40	-	-

\*Land and air transport

Source: Observatoire des Sciences et des Techniques (1992), pp. 142-143.

This table provides a breakdown of the “scientific intensity” of patents, by industrial and technological activity of their applicants. References are to SCI scientific papers cited by one or more patents. “Pharmaceuticals is the activity whose technology is nearest to science, followed by chemicals, instruments and electronics. In contrast, transport technology makes practically no reference to scientific publications. In terms of the time needed to move from science to technology, electronics and pharmaceuticals are set apart by their particular speed; in contrast, chemicals tends to rely more on older science” (as translated from OST, 1992, Chapter 3).

**Example No. 21.**  
**Distribution by scientific discipline of scientific citations in world patent applications in  
the United States (averages, 1987-88)**

Scientific discipline	Citations by patents (%)
Clinical medicine	20.0
Biomedical research	20.5
Animal and vegetable biology	2.5
Chemistry	23.0
Physics	18.0
Earth and space sciences	0.0
Engineering sciences	15.5
Mathematics	0.0

Source: Observatoire des Sciences et des Techniques (1992), pp. 142-143.

This table (based on the same statistics as the previous example) shows the various disciplines' shares in citations of scientific articles by patents. "The distribution by scientific discipline of articles cited by patents shows that clinical medicine, biomedical research, chemistry, physics and engineering sciences share the citations almost equally. Animal and vegetable biology, earth and space sciences and mathematics are disciplines to which patents make little or no reference" (as translated from OST, 1992, Chapter 3).

**Example No. 22.**  
**Production and utilisation of science in connection with patents**  
**(1987-88 average) in worldwide weighting**

<b>Production</b>					
<b>Science</b>	<b>Science with technological potential</b>	<b>Science utilised</b>	<b>Patents</b>	<b>Science citation</b>	
France	4.9	4.8	3.3	3.7	2.9
Germany	6.0	6.6	3.8	10.3	5.2
United Kingdom	8.9	8.1	7.5	3.6	3.8

*Source:* Observatoire des Sciences et des Techniques (1992), pp. 142-143.

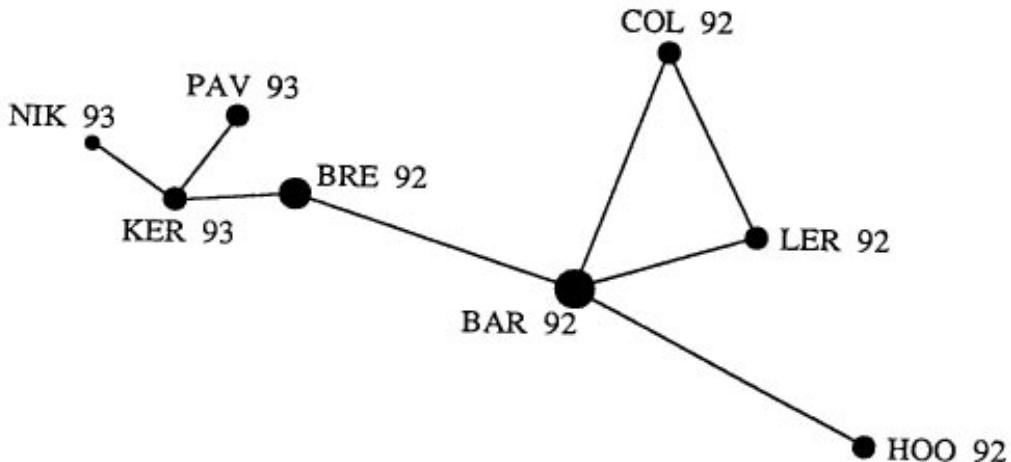
Based also on the same statistics, this table shows the following parameters (indicators) for France, Germany and the United Kingdom (all expressed as a percentage of the world total):

- scientific production: the three countries' respective shares of scientific articles (all disciplines combined);
- scientific production with technological potential (= science likely to be cited by patents): the countries' average worldwide shares of scientific articles, by discipline, weighted by their technological potential;
- scientific production "utilised" by patents: the countries' worldwide shares of scientific articles cited by aggregate patent applications (all countries combined);
- patent production: the countries' world share of patents applied for in the United States;
- utilisation of science in patents: the countries' world shares of citations, by patents, of scientific articles (number of citations of scientific articles in each country's patents, as a percentage of citations by patents worldwide).

"German scientific production with technological potential is only slightly less than that of the United Kingdom (6.6 per cent of the world total, versus 8.1 per cent), but it is much less 'utilised' (cited by patents) than British science (3.8 per cent of the world total, versus 7.5 per cent). French scientific production is almost as 'utilised' for world technological creation as German science (3.3 per cent versus 3.8 per cent).

"It is Germany, however, that most 'utilises' science in absolute terms for its technology (5.2 per cent); this reflects Germany's substantial technological weight (10.3 per cent), which is nearly three times that of the United Kingdom or France." (as translated from OST, 1992, Chapter 3).

**Example No. 23.**  
**Clusters of co-citations in 1992 et 1993**  
**in the field of “combinative chemistry”**



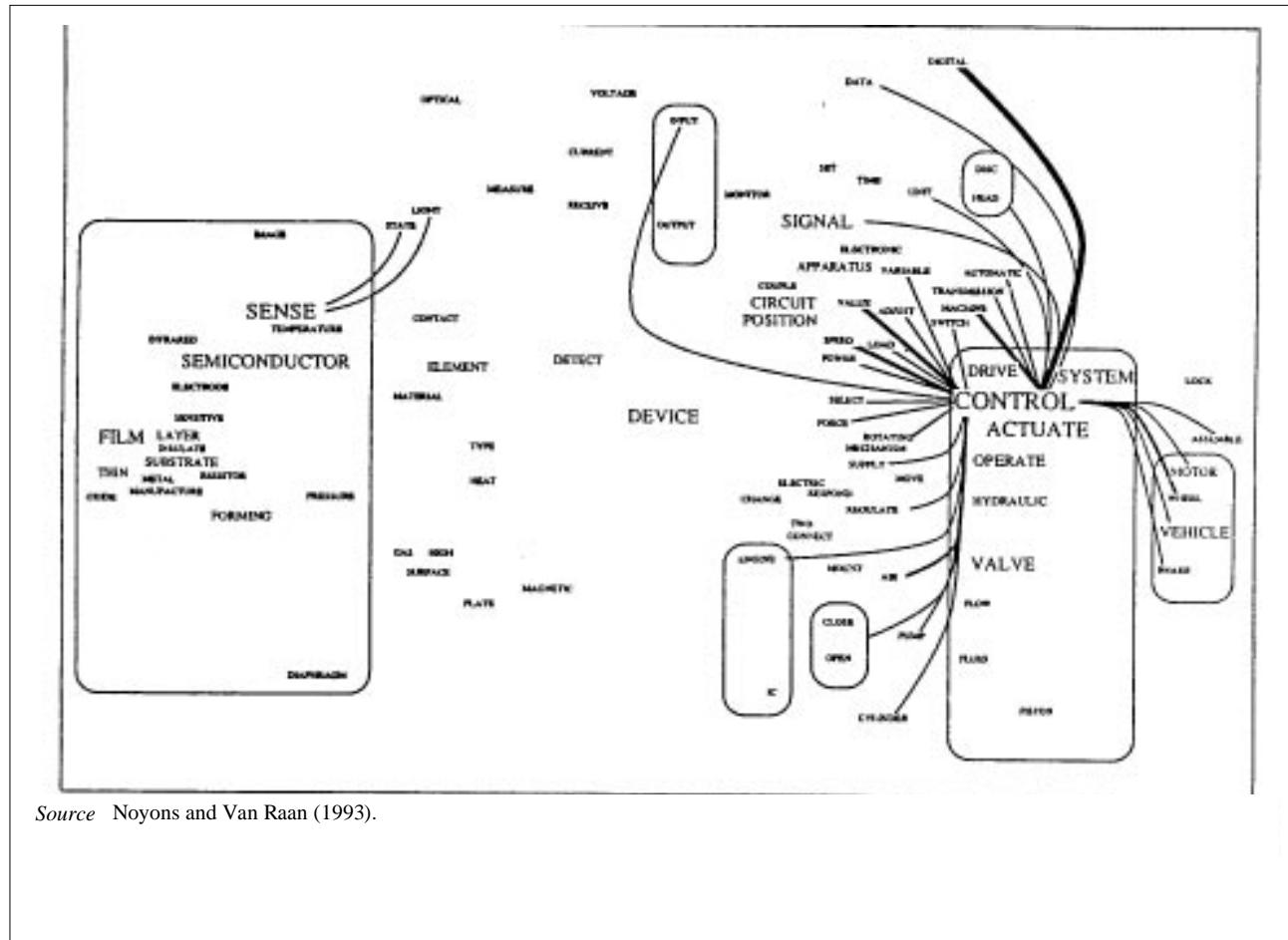
BAR : BARBAS C.F.      COL : COLLET T.A.      LER : LERNER R.A.      HOO : HOOGENBOOM H.R.  
BRE : BRENNER S.      KER : KERR J.M.      PAV : PAVIA M.R.      NIK : NIKOLAEV V.

Source: Data specially provided to Yoshiko Okubo by Henry Small of the Institute for Scientific Information.

This diagram shows an example of co-citations that is taken from an ISI study on combinative chemistry – a promising field for the development of new drugs – and its biological applications. The abbreviations refer to researchers working in this field. The lines represent linkages between these people, which are shown by co-citations. The thickness of the points of reference reflect the intensity of the connections.

The cluster of co-cited publications shows the existence of a number of groups of scientists sharing the same interests and the same references, as identified by connected key-words (see Chapter 5). In 1992, an initial cluster formed around one of the authors, Barbas, who was then at the centre of this research. In 1993, the cluster grew, with three additional articles around Kerr, and the subject shifted towards another sub-group of the same research (“peptide groups”, a theme identified by the titles of papers).

## **Example No. 24.** **Map of the “technology field” in optomechatronics**



*Source* Noyons and Van Raan (1993).

This map, prepared using a matrix-based technique, shows the “technological” side of work in the field of optomechatronics in 1991. It is based on an analysis of the frequencies of co-occurrence, in technology patents, of key words specific to the field. Associated words are mapped according to their mutual relationships, including “hidden links” which would be difficult to detect in other formats. The relative distances between research topics, defined as key words (whose size on the map denotes their frequency), reflect their cognitive relationships.

**Example No. 25.**  
**Matrix of international collaboration, 1981-86: 98 countries and eight fields of science**

	<b>MAT</b>	<b>PHY</b>	<b>CHM</b>	<b>ENT</b>	<b>EAS</b>	<b>BIO</b>	<b>BIM</b>	<b>CLI</b>	<b>Total</b>
United States	4 077	16 579	6 952	5 246	8 079	5 949	15 686	20 698	84 216
United Kingdom	1 168	7 334	4 099	1 860	3 601	2 492	6 639	9 158	36 763
Germany	962	9 964	3 540	1 446	2 935	1 790	5 687	5 714	32 219
France	931	8 387	3 433	898	2 263	1 280	4 530	5 287	27 521
Canada	1 335	3 376	2 103	1 614	2 119	2 001	3 501	4 967	21 129
Italy	410	4 70	1 699	402	1 268	511	2 160	3 344	14 652
Switzerland	193	5 880	984	426	508	374	2 150	3 616	14 213
World	13 758	89 429	41 275	20 255	32 103	28 282	64 589	92 488	382 179

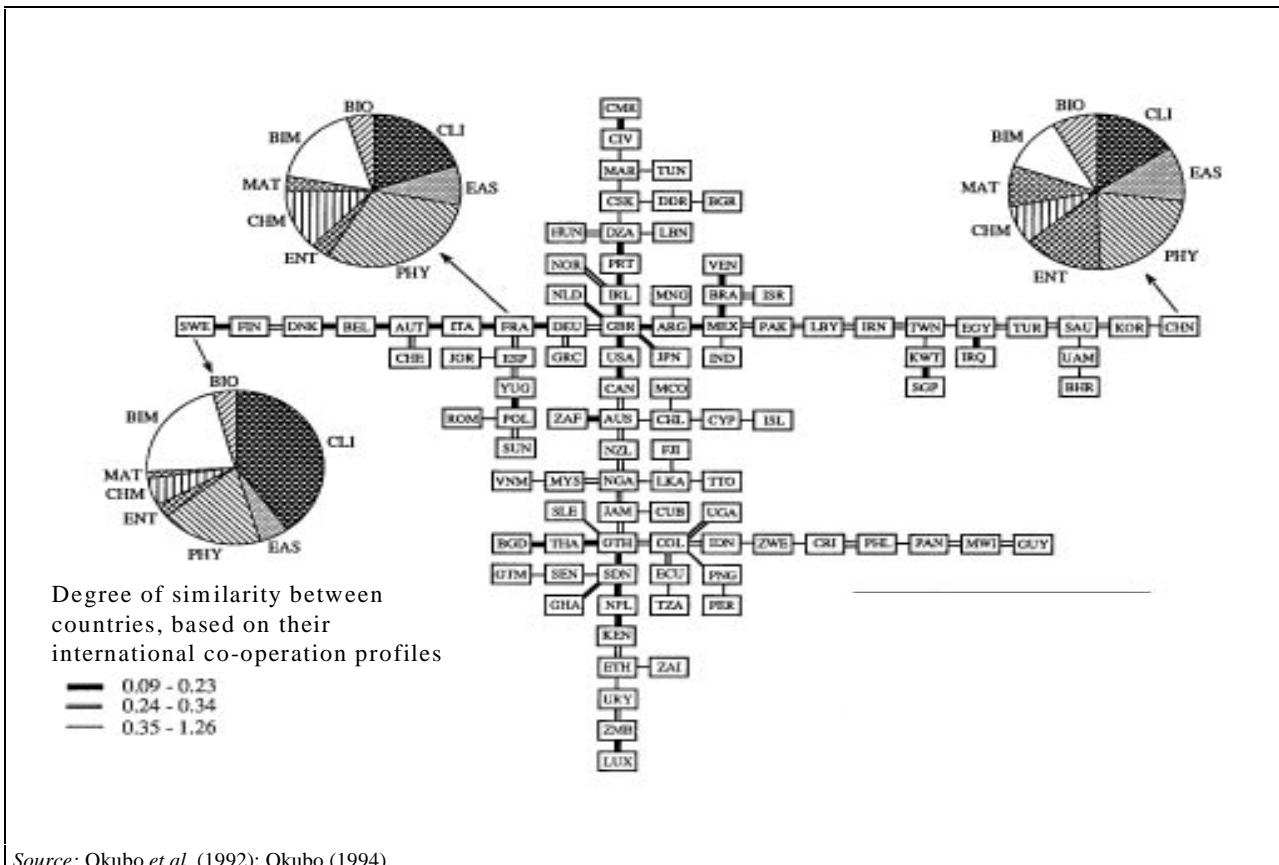
Source: See Examples 25A, B

MAT (mathematics), PHY (physics), CHM (chemistry), ENT (engineering-technology), EAS (Earth-space), BIO (basic and applied biology), BIM (biomedicine), CLI (clinical medicine).

Examples 25 and 25A and B highlight the different behaviour of the various countries. The table for Example 25 shows the number of international links, as measured in terms of co-authorship, for the seven leading countries in international scientific collaboration, as well as the world total – all broken down into the eight main fields of science (for the abbreviations of country names, see Annex I). It is therefore a multidimensional analysis comprising 98-times-eight dimensions.

Using this table it is possible, for each country, to construct a pie chart representing these international links by field of science. Since the proportions vary by country, each country has its own profile. On this basis, countries are grouped together, on the maps, in clusters of like profiles.

**Example No. 25A.**  
**“Minimum spanning tree”**  
**Similarities between countries based on their patterns of international collaboration,  
 1981-86**



Source: Okubo *et al.* (1992); Okubo (1994).

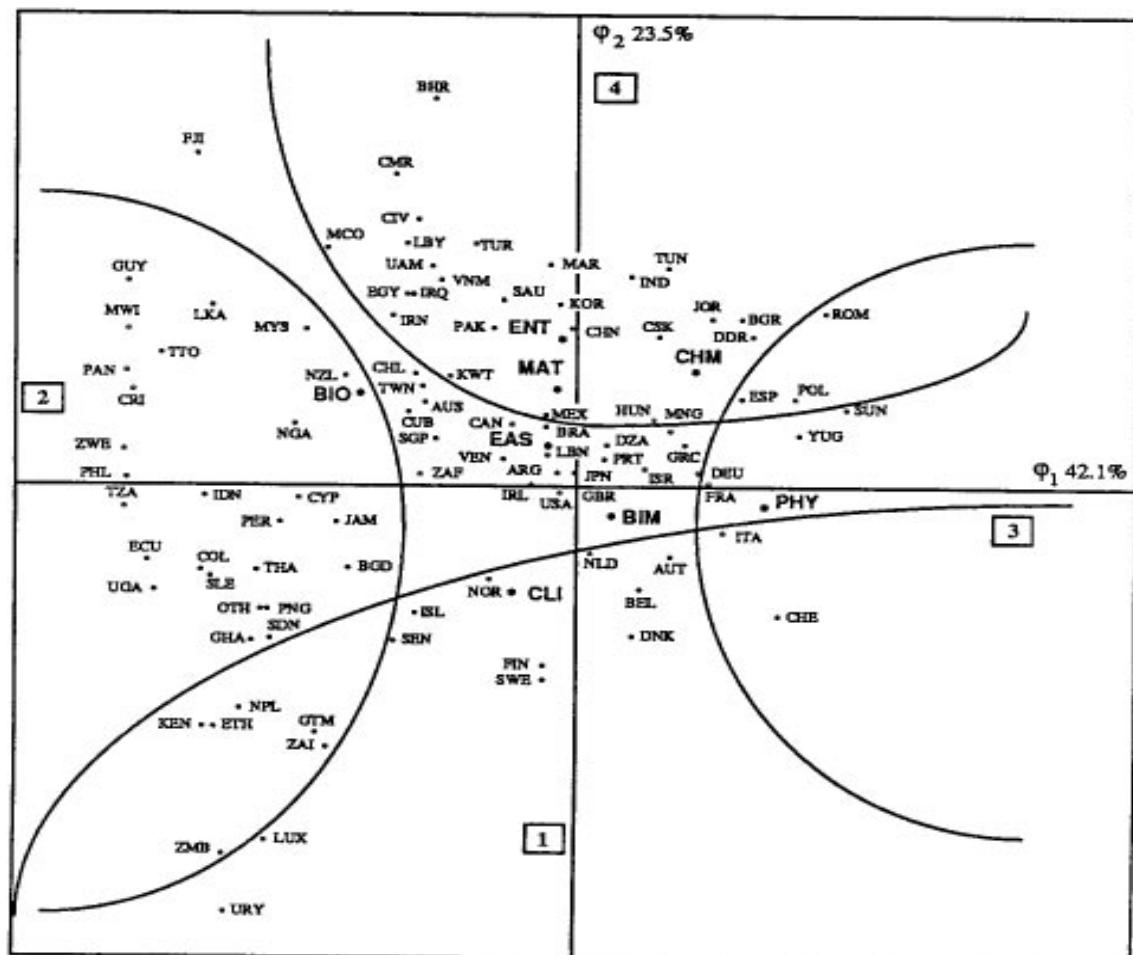
This “map” is generated using the classification technique known as the “minimum spanning tree” (MST). It classifies the 98 countries in Example 25 with respect to the “world model” according to the similarity of the profiles of their international links in the eight fields of science over the period. Each rectangle represents a country and replaces the corresponding pie charts – for obvious technical reasons. The “degrees of similarity” between countries are illustrated by the thickness of the lines depicting the links.

Three other charts illustrate the profiles of the two “extreme countries” along the longest horizontal line (Sweden and China), whose co-operative ties are vastly different, as well as the profile of France. The first two are fairly dissimilar from the “world model”, which, lying near the centre of the tree, is heavily influenced by the co-operative structures of the leading scientific countries (including Germany, the United States, France and the United Kingdom). Similarly, there are sharp differences in co-operative links between countries along the vertical axis, along which lie countries from Cameroon (at the top) to Luxembourg (at the bottom).

All of the countries concentrate their international exchanges in specific fields, irrespective of their size or efforts to adhere to the centre of gravity of scientific activity. Along the first main branch, the European countries (tending to specialise in clinical medicine) line up on the left, up to the United Kingdom (GBR), whereas the Arab and Islamic countries, along with Asian countries, line up on the right (all fairly oriented towards co-operation in the fields of physics and engineering/technology). Using this MST, a large number of historical, cultural, economic and political factors may be cited.

### **Example No. 25 B.**

## **Relative positions of the 98 countries with regard to international co-operation in the eight fields of science, 1981-86**



## Correspondence factorial analysis

- (1) Country highly active in clinical medicine.
  - (2) Country highly active in basic and applied biology.
  - (3) Country highly active in physics.
  - (4) Country highly active in engineering, mathematics and chemistry.

MAT (mathematics), PHY (physics), CHM (chemistry), ENT (engineering-technology), EAS (earth-space), BIO (basic and applied biology), BIM (biomedicine), CLI (clinical medicine).

The phi 1, 2, and 3 axes represent all of the information contained in the table analysed. The first two axes (phi 1 and 2) convey most of the information and are most representative of an entire matrix.

Source: Okubo et al. (1992); Okubo (1994)

Source: Okubo et al. (1992); Okubo (1994).

This map uses correspondence factorial analysis to depict international collaboration and its specificities, for the same 98 countries and eight fields of science from Example 25, by positioning the elements (here, the countries and fields of science).

The “keys” for interpreting this map are as follows:

- The closer together the countries are on the chart, the more similar are their profiles of collaboration with the group of all countries in the eight fields of science.
- The closer a country is to a field, the greater is its collaboration in the field, as compared with that of the other countries.
- The closer a country is to the centre of the diagram, the more balanced is its collaboration in the various fields of science.

An example to illustrate the use of these “keys” is the group located near the centre of the map. This group, which among other countries includes Japan (JPN) and the United States (USA), exhibits similar patterns of co-operation, engaging in more or less equivalent international collaboration in each of the four fields indicated.

The group located in the upper left-hand frame (quadrant 2), which includes, for example, Australia (AUS), Canada (CAN) and New Zealand (NZL), concentrates its collaboration in the four fields indicated (BIO, EAS, MAT and ENT), with New Zealand showing a high degree of specialisation, especially in biology.

The Nordic countries – Iceland (ISL), Norway (NOR), Finland (FIN) and Sweden (SWE) – concentrate on clinical medicine (CLI) (quadrant 1), while at the same remaining very close to Denmark (DNK), Belgium (BEL) and the Netherlands (NLD), which are nonetheless closer to the biomedical field (BIM) of quadrant 3.

## NOTES

1. This list follows that of Harada (1986, pp. 139-143), who nonetheless presented six categories rather than the four described here.
2. Published by the Information Science and Scientometrics Research Unit (ISSRU), Library of the Hungarian Academy of Sciences (Akadémiai Kiado és Nyomda Vallalat), Budapest, Hungary. The first issue came out in September 1978.
3. For patent databases, see also OECD, 1994, Section 3.3.
4. Physikalische Berichte, a database produced by the German firm Deutsche Physikalische Gesellschaft.
5. For example, a comparative study by Moed (1989) proposes an assessment of mathematics in Italy, measured in two different ways. The “fixed” base indicates a much lower level of activity than the measurement from the variable set. This discrepancy is due to the appearance of two mathematical journals, *Bollettino dell’Unione Matematica Italiana* and *Annali di Matematica Pura e Applicata*, in which Italian scientists are prominent and which were incorporated into SCI during the freeze period of the fixed journal set.

## REFERENCES

- ANDERSON, J., P.M.D. COLLINS, J. IRVINE, P.A. ISARD, B.R. MARTIN, F. NARIN, and K. STEVENS (1989), "On-line Approaches to Measuring National Scientific Output: A Cautionary Tale", *Science and Public Policy*, Vol. 15 (3), pp. 153-161.
- BRADFORD, S.C. (1950), *Documentation*, Public Affairs Press, Washington, DC.
- BRAUN, T., W. GLÄNZEL, and A. SCHUBER (1991), "The Bibliometric Assessment of UK Scientific Performance: Some Comments on Martin's 'Reply'", *Scientometrics*, Vol. 20 (2), pp. 359-362.
- BUSH, V. (1960), "Science and the Public Welfare", in *Science: The Endless Frontier*, National Science Foundation (Reprint), Washington, DC, pp. 17-27.
- CALLON, M., J.P. COURTIAL and H. PENAN (1993), *La scientométrie*, Presses universitaires de France, Paris.
- CALLON, M. and L. LEYDESDORFF (1987), "La recherche française est-elle en bonne santé?", *La Recherche*, Vol. 18 (186), pp. 412-419.
- CHAUVIN, R. (1991), "Évaluer, évaluer", *La Recherche*, Vol. 22, p. 782.
- CHELIMSKY, E. (1991), "On the Social Science Contribution to Governmental Decision-Making", *Science*, Vol. 254, pp. 226-231.
- COLE, J. and N.B. EALES (1917), "The History of Comparative Anatomy: A Statistical Analysis of the Literature", *Science Progress*, Vol. 11 (4), pp. 578-596.
- CUNNINGHAM, E.R. (1935), "The Present Status of the Publication of Literature in the Medical and Biological Sciences", *Medical Library Association Bulletin*, Vol. 24, pp. 64-81.
- FRAME, J.D. (1977), "Mainstream Research in Latin America and the Caribbean", *Interciencia*, Vol. 2 (3), pp. 143-148.
- GARFIELD, E. (1968), *World Brain or Memex? Mechanical and Intellectual Requirements for Universal Bibliographic Control, The Foundations of Access to Knowledge*, Syracuse University Press, New York.
- GARFIELD, E. (1972), "The Design and Production of a Citation Index", in *Citation Indexing-Its Theory and Application in Science, Technology and Humanities*, John Wiley & Sons, New York, pp. 19-36.
- GARFIELD, E. (1975), "Introduction", in *Journal Citation Reports: Science Citation Index*, Vol. 9 (3).
- GARFIELD, E. (1979a), "How Do We Select Journals for Current Contents?", *Current Contents*, 5 November.
- GARFIELD, E. (1979b), "Is Citation Analysis a Legitimate Evaluation Tool?", *Scientometrics*, Vol. 1, pp. 359-375.
- GARFIELD, E. (1988), "French Research: Citation Analysis Indicates Trends are More than Just a Slip of the Tongue", *Current Contents*, Vol. 23, p. 9.
- GLÄNZEL, W. and U. SCHOEPPFLIN (1994), "Little Scientometrics, Big Scientometrics . . . and Beyond", *Scientometrics*, Vol. 30 (2-3), p. 375.
- HARADA, M. (1986), "La tendance de la bibliométrie", *Operations Research*, March, pp. 139-143.
- HIRASAWA, R. (1995), "Integrative Management of Technology - Analysis and Its Implications", R&D Dynamics Network Meeting in Stockholm, 7-9 November.
- HULME, E.W. (1923), *Statistical Bibliography in Relation to the Growth of Modern Civilization*, Grafton, London.

- INSTITUTE FOR SCIENTIFIC INFORMATION (1981), *Science Citation Index 1981 Guide & Lists of Source Publications*, Institute for Scientific Information, Inc., Philadelphia.
- IRVINE, J. and B.R. MARTIN (1980), "The Economic Effects of Big Science: The Case of Radio Astronomy", Science & Technology Indicators Conference, OECD, Paris.
- KATZ, J.S. and D.M. HICKS (1996), "A Systemic View of British Science", *Scientometrics* 35, No. 1, pp. 133-154.
- KEALEY, T. (1991), "Government-Funded Academic Science is a Consumer Good, Not a Producer Good: A Comparative Reassessment of Britain's Scientific and Technological Achievements since 1794 and a Comment on the Bibliometry of B. Martin and J. Irvine", *Scientometrics*, Vol. 20 (2), pp. 369-394.
- KOBAYASHI, S. (1987), "Internationalisation of Japanese Journals and their Contribution to Basic Science", *Daigaku Kenkyu*, Vol. 1, pp. 57-76.
- LEVEN, O. (1982), *Report of the National Survey Committee on Biochemistry*, Staatsuitgeverij, The Hague.
- LEYDESDORFF, L. (1991), "On the 'Scientometric Decline' of British Science. One Additional Graph in Response to Ben Martin", *Scientometrics*, Vol. 20 (2), pp. 363-368.
- LEYDESDORFF, L. (1995), *The Challenge of Scientometrics: The Development, Measurement, and Self-organisation of Scientific Communication*, DSWO Press, Leiden.
- LOTKA, A.J. (1926), "The Frequency Distribution of Scientific Productivity", *Journal of the Washington Academy of Science*, Vol. 16 (12), pp. 317-323.
- MAGRI, M.H. and A. SOLARI (1996), "The SCI Journal Citation Reports: A Potential Tool for Studying Journals?", *Scientometrics*, Vol. 35, pp. 93-117.
- MARTIN, B.R. (1991), "The Bibliometric Assessment of UK Scientific Performance. A Reply to Braun, Glänzel and Schubert", *Scientometrics*, Vol. 20 (2), pp. 333-357.
- MARTIN, B.R. and J. IRVINE (1984), "CERN: Past Performance and Future Prospects", *Research Policy*, Vol. 13, pp. 311-342.
- MARTIN, B.R. and J. IRVINE (1985) "Basic Research in the East and West: A Comparison of the Scientific Performance of High-Energy Physics Accelerators", *Social Studies of Science*, Vol. 15, pp. 293-394.
- MERTON, R. K. (1957b), "Social and Democratic Social Structure", in *Social Theory and Social Structure*, Free Press, New York, pp. 550-561.
- MERTON, R.K. (1957a), "Priorities in Scientific Discovery", *American Sociological Review*, Vol. 22, p. 635.
- MINISTRY OF EDUCATION, SCIENCE AND CULTURE (1987), "International Comparative Study of the Number of Scientific Articles", Japan.
- MIQUEL, J.F. and J.C. DORÉ (1981), "Research on New Drugs: A System Analysis Approach", in *Drugs of Today XVII* (11), pp. 487-501.
- MIQUEL, J.F. and Y. OKUBO (1990), "Indicators to Measure Internationalization of Science", *Consequences of the Technology Economy Programme for the Development of Indicators*, OECD, Paris.
- MIQUEL, J.F. and Y. OKUBO (1994), "Structure of International Collaboration in Science-Part II: Comparisons of Profiles in Countries Using a Link Indicator", *Scientometrics* 29 (2), pp. 271-297.
- MOED, H.F. (1989), "New Developments in Online Bibliometric Analysis", in *Science and Technology Indicators*, DSWO Press, Leiden, pp. 129-146.
- MOED, H.F., W.J.M. BURGER, J.G. FRANKFORT, and A.F.J. VAN RAAN (1983), "On the Measurement of Research Performance: The Use of Bibliometric Indicators", Leiden University, Leiden.
- NATIONAL SCIENCE BOARD (1989), "Public Attitude towards Science in General", in *Science and Engineering Indicators 1989*, National Science Foundation, Washington, DC.

- NATIONAL SCIENCE BOARD (1989), *Science and Engineering Indicators 1989*, National Science Foundation, Washington, DC.
- NATIONAL SCIENCE FOUNDATION (1972), *Science and Engineering Indicators 1972*, National Science Foundation, Washington, DC.
- NATIONAL SCIENCE FOUNDATION (1989), *Science & Engineering Indicators 1989*, National Science Foundation, Washington, DC.
- NEDERLANDS OBSERVATORIUM VAN WETENSCHAP EN TECHNOLOGIE (CENTRUM VOOR WETENSCHAPS- EN TECHNOLOGIE-STUDIES (1994), *Wetenschaps- en Technologie-Indicatoren 1994*, Rijksuniversiteit, Leiden.
- NIWA, F. and H. TOMIZAWA (1995), "Development of General Indicators of Science and Technology", in *Proceedings of the Fourth International Conference on Science and Technology Indicators*, Antwerp.
- NOYONS, E.C.M. and A.F.J. VAN RAAN (1993), "Science Base of Technology - Bibliometric Mapping as a Tool for National Science and Technology Policy: Part II: Optomechatronics", CWTS Report 93-07, Leiden University.
- NOYONS, E.C.M. and A.F.J. VAN RAAN (1995), "Mapping the Development of Neural Network Research: Structuring the Dynamics of Neural Network Research and an Estimation of German Activity", CWTS Report 95-06, Centrum voor Wetenschaps- en Technologie-Studies.
- OECD (1989), "Bibliometric Analyses", in *Measurement of R&D Output, The Measurement of Scientific and Technical Activities, R&D Statistics and Output Measurement in the Higher Education Sector, "Frascati Manual"* supplement, Paris.
- OECD (1994), "Using Patent Data as Science and Technology Indicators - Patent Manual 1994", Paris.
- OKUBO, Y., J.F. MIQUEL, L. FRIGOLETTA and J.C. DORÉ (1992), "Structure of International Collaboration in Science; Typology of Countries through Multivariate Techniques Using a Link Indicator", *Scientometrics* 25, Vol.2, pp. 321-351.
- OKUBO, Y. (1993), "International Co-operative Network in Drug Innovation International R&D Network", Kyoto Conference, May 1993.
- OKUBO, Y. (1994), "L'Internationalisation de la Science : Création d'indicateurs bibliométriques pour une mise à jour de l'activité scientifique internationale du Japon", doctoral dissertation.
- OBSERVATOIRE DES SCIENCES ET DES TECHNIQUES (1992), "Science & Technologie Indicateurs", *Economica*, Paris.
- OTSU, K. (1983), "A Bibliometric Study of Japanese Science and Social Science Publications", *Library and Information Science*, Vol. 21, pp. 19-27.
- OURISSON, G. (1991), "Une évaluation pertinente", *La Recherche*, Vol. 22, pp. 1209-1210.
- PENDLEBURY, D. (1991), "Research Papers: Who's Uncited Now?", *Science*, Vol. 251, p. 25.
- PRICE, D. (1963), *Little Science, Big Science*, Columbia University Press, New York.
- PRICE, D. (1964), *The Science of Science*, London Souvenir Press, London.
- PRICE, D. (1965), "Networks of Scientific Papers", *Science*, Vol. 149, pp. 510-515.
- PRICE, D. (1969), "Policies for Science?", *Melbourne Journal of Politics*, Vol. 2, pp. 1-8.
- PRITCHARD, J. (1969), "Statistical-Bibliography or Bibliometrics?", *Journal of Documentation*, Vol. 25 (4), pp. 348-349.
- RAPPA, M.A. (1989), "Bibliometric Methods for Monitoring the Emergence of New Technologies", Report No. WP3049-89-BPS, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- ROYAL SOCIETY (1986), *Evaluation of National Performance in Basic Research*, study by the Royal Society research service for the Advisory Board for the Research Councils, The Royal Society, London.
- SEGLEN, Per O. (1992), "The Skewness of Science", *Journal of the American Society for Information Science*, Vol. 43, pp. 628-638.

- SIVERTSEN, G. (1991), "Should a New Database for International Comparison be More Restricted in Journal Coverage?", in *Science and Technology in a Policy Context*, Acts of the EC-Leiden Colloquium on Science and Technology Indicators, DSWO Press, Leiden.
- THE ECONOMIST (1992), "Pity the Typists", 18 January, p. 87.
- VAN RAAN, A.F.J. and R.J.W. TIJSSEN (1990), "An Overview of Quantitative Science and Technology Indicators Based on Bibliometric Methods", Technology Economy Programme for the Development of Indicators, OECD, Paris.
- WADE, N. (1975), "Citation Analysis: A New Tool for Science Administrators, *Science*, Vol. 188, pp. 429-432.

## ADDITIONAL REFERENCES FOR FURTHER READING

- BONITZ, M., E. BRUCKNER and A. SCHARNHORST (1993), "The Science Strategy Index", *Scientometrics* 26 (1), pp. 37-50.
- BRAAM, R.F., H.F. MOED and A.F.J. VAN RAAN (1991), "Mapping of Science by Combining Co-citation and Word Analysis. I. Structural Aspects", *Journal of the American Society for Information Science* 42, pp. 233-251.
- BRAAM, R.F., H.F. MOED and A.F.J. VAN RAAN (1991), "Mapping of Science by Combining Co-citation and Word Analysis. II. Dynamical Aspects", *Journal of the American Society for Information Science* 42, pp. 252-266.
- BOURKE, P. and L. BUTLER (1996), "Standard Issues in a National Bibliometric Database: The Australian Case", *Scientometrics* 35 (2), pp. 199-208.
- CALLON, M., J. LAW J. and A. RIP (eds) (1986), *Mapping the Dynamics of Science and Technology*, Macmillan, London.
- CARPENTER, M. (1979), *International Science Indicators – Development of Indicators of International Scientific Activity using the Science Citation Index*, Report No. SRS77-22770, National Science Foundation, Washington, DC.
- CARPENTER, M. and F. NARIN (1981), "The Adequacy of the Science Citation Index (SCI) as an Indicator of International Scientific Activity", *Journal of the American Society of Information Science* 32, pp. 430-439.
- CHUBIN, D. and K. STUDER (1979), "Knowledge and Structures of Scientific Growth: Measurement of a Cancer Problem Domain", *Scientometrics* 1, pp. 171-193.
- CHUBIN, D. (1986), "Non-economic Quantitative Measures: The 'Output' of Science", in *Research Funding as an Investment. Can we Measure the Returns? A Technical Memorandum*, Office of Technology Assessment, Washington, DC, pp. 13-44.
- COZZENS, S.E. (1989), "What do Citations Count? The Rhetoric First Model", *Scientometrics* 15 (5-6), pp. 437-447.
- COZZENS, S.E. (1989), *Literature-Based Data in Research Evaluation: A Manager's Guide to Bibliometrics*, Department of Science and Technology Studies, Rensselaer Polytechnic Institute Troy, New York.
- COZZENS, S.E. (1995), "U.S. Research Assessment: Recent Developments", *Scientometrics* 34 (3), pp. 351-362.
- EDGE, D. (1979), "Quantitative Measures of Communication in Science: A Critical Review", *History of Science* 17, pp. 102-134.
- ELKANA, K., L. LEDERBERG, R.K. MERTON, A. THACKRAY and H. ZUCKERMAN (1978), *Toward a Metric of Science: The Advent of Science Indicators*, John Wiley, New York.
- GARFIELD, E. (1972), *Citation Indexing – Its Theory and Application in Science, Technology and Humanities*, John Wiley & Sons, New York.
- GLANZEL, W. (1996), "The Need for Standards in Bibliometric Research and Technology", *Scientometrics* 35 (2), pp. 167-176.
- GLANZEL, W. (1996), "A Bibliometric Approach to Social Sciences, National Research Performances in Six Selected Social Science Areas, 1990-1992", *Scientometrics* 35 (3), pp. 291-308.
- GOMEZ, I., M. BORDONS, M.T. FERNANDEZ and A. MENDEZ (1996), "Coping with the Problem of Subject Classification Diversity", *Scientometrics* 35 (2), pp. 223-236.
- HJORTGAARD C. and P. INGWERSEN (1996), "Online Citation Analysis. A Methodological Approach", *Scientometrics* 37 (1), pp. 39-62.

- HOFER GEE, H. and F. NARIN, *An Analysis of Research publications Supported by NIH, 1973-1980: National Institute of Health and NIH Intramural Program*, NIH Program Evaluation Report, pp. 16-17.
- KATZ, J.S. and D.M. HICKS (1996), "A Systemic View of British Science", *Scientometrics* 35 (1), pp. 133-154.
- KATZ, J.S. (1996), "Bibliometric Standards: Personal Experience and Lessons Learned", *Scientometrics* 35 (2), pp. 193-199.
- LEYDESDORFF, L. (1987), "Various Methods for the Mapping of Science", *Scientometrics* 11, pp. 295-324.
- LEYDESDORFF, L. and S.E. COZZENS (1993), "The Delineation of Specialities in Terms of Journals Using the Dynamic Journal Set of the SCI", *Scientometrics* 26 (1), pp. 135-156.
- LUUKKONEN, T. (1989), "Publish in a Visible Journal or Perish? Assessing Citation Performance of Nordic Cancer Research", *Scientometrics* 15 (5-6), pp. 349-367.
- MACROBERTS, M.H. and B.R. MACROBERTS (1986), "Quantitative Measures of Communication in Science: A Study of the Formal Level", *Social Studies of Science* 16, pp. 151-172.
- MARTIN, B.R. and J. IRVINE (1983), "Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy", *Research Policy* 12, pp. 61-90.
- MARTIN, B.R. (1996), "The Use of Multiple Indicators in the Assessment of Basic Research", *Scientometrics* 36 (3), pp. 343-362.
- MELIN, G. and O. PERSSON (1996), "Studying Research Collaboration Using Co-authorships", *Scientometrics* 36 (3), pp. 363-378.
- NARIN, F. (1976), *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*, National Science Foundation, Washington, DC.
- NARIN, F. and S. KIMBERLY HAMILTON (1996), "Bibliometric Performance Measures", *Scientometrics* 36 (3), pp. 293-310.
- NOMA, E. (1982), "An Improved Method for Analysing Square Scientometric Transaction Matrices", *Scientometrics* 4, pp. 297-316.
- PAVITT, K. (1985), Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems, *Scientometrics* 7, pp. 77-99.
- PORTER, A.L. and J.D. ROESSNER (1991), *Indicators of National Competitiveness in High Technology Industries*, Final Report to the Science Indicators Studies Group, National Science Foundation, 2 volumes, Georgia Institute of Technology, Atlanta.
- PRIM, R.C. (1957), "Shortest Connection Networks and Some Generalizations", *Bell System Technology Journal* 36, p. 1389.
- RIVAS, A.L., D. DESHLER, R.D. COLLE, R.N. GONZALEZ and F.W. QUIMBY (1996), "Indicators of Disciplinary Differentiation: Interdisciplinary Linkages and Adoption Rate of Biological Innovations", *Scientometrics* 37 (1), pp. 63-86.
- SEGLEN, PER O. (1996), "Quantification of Scientific Article Contents", *Scientometrics* 35 (3), pp. 355-366.
- SCHUBERT, A., W. GLÄNZEL and T. BRAUN (1989), "Scientometric Datafiles: A Comprehensive Set of Indicators on 2 649 Journals and 96 Countries in All Major Science Fields and Subfields 1981-1985", *Scientometrics* 16, 3-478.
- SCHUBERT, A. and T. BRAUN (1990), "International Collaboration in the Sciences, 1981-1985," *Scientometrics*, 19 (1-2), pp. 3-10.
- SCHUBERT, A. and T. BRAUN (1996), "Cross-Field Normalization of Scientometric Indicators", *Scientometrics* 36 (3), pp. 311-324.
- SMALL, H. (1973), "Co-citation in the Scientific Literature: A New Measure of the Relationship between Publications", *Journal of the American Society for Information Science* 24, pp. 265-269.

- SMALL, H. and E. GARFIELD (1985), "The Geography of Science: Disciplinary and National Mappings", *Journal of Information Science* 11, pp. 147-159.
- SMALL, H. and E. GREENLEE (1985), "Clustering the Science Citation Index using Co-citations. I: A Comparison of Methods", *Scientometrics* 7, pp. 391-409.
- SMALL, H. (1994), "A SCI-map, Case Study: Building a Map of AIDS Research", *Scientometrics* 30, pp. 229-241.
- VAN RAAN, A.F.J. (ed.) (1988), *Handbook of Quantitative Studies of Science and Technology*, Elsevier.
- VAN RAAN, A.F.J. (1996), "Advanced Bibliometric Methods as Quantitative Core of Peer Review Based Evaluation and Foresight Exercises", *Scientometrics* 36 (3), pp. 397-420.
- VINKLER, P. (1996), "Relationships between the Rate of Scientific Development and Citations. The Chance for Citedness Models", *Scientometrics* 35 (3), pp. 373-386.
- VINKLER, P. (1993), "Research Contribution, Authorship and Team Cooperativeness", *Scientometrics* 26 (1), pp. 213-229.

**ANNEX I.**  
**INDEX OF ISO CODES**

ARG	Argentina	KWT	Kuwait
AUS	Australia	LBN	Lebanon
AUT	Austria	LBY	Libyan Arab Jamahiriya
BEL	Belgium	LKS	Sri Lanka
BGC	Bangladesh	LUX	Luxembourg
BGR	Bulgaria	MAR	Morocco
BHR	Bahrain	MCO	Monaco
BRA	Brazil	MEX	Mexico
CAN	Canada	MLI	Mali
CHE	Switzerland	MNG	Mongolia
CHL	Chile	MWI	Malawi
CHN	China	MYS	Malaysia
CIV	Côte d'Ivoire	NGA	Nigeria
CMR	Cameroon	NLD	Netherlands
COL	Colombia	NOR	Norway
CRI	Costa Rica	NPL	Nepal
CSK	Czechoslovakia	NZL	New Zealand
CUB	Cuba	PAK	Pakistan
CYP	Cyprus	PAN	Panama
DDR	German Democratic Republic	PER	Peru
DEU	Germany, Federal Republic of	PHL	Philippines
DNK	Denmark	PNG	Papua New Guinea
DZA	Algeria	POL	Poland
ECU	Ecuador	PRT	Portugal
EGY	Egypt	ROM	Romania
ESP	Spain	SAU	Saudi Arabia
ETH	Ethiopia	SDN	Sudan
FIN	Finland	SEN	Senegal
FJI	Fiji	SGP	Singapore
FRA	France	SLE	Sierra Leone
GBR	United Kingdom	SUN	USSR
GHA	Ghana	SWE	Sweden
GRC	Greece	THA	Thailand
GTM	Guatemala	TTO	Trinidad and Tobago
GUY	Guyana	TUN	Tunisia
HUN	Hungary	TUR	Turkey
IDN	Indonesia	TWN	Taiwan, Province of China
IND	India	TZA	Tanzania, United Republic of
IRN	Iran	UGA	Uganda
IRL	Ireland	URY	Uruguay
IRQ	Iraq	UAM	United Arab Emirates
ISL	Iceland	USA	United States
ISR	Israel	VEN	Venezuela
ITA	Italy	VNM	Viet Nam
JAM	Jamaica	YUG	Yugoslavia
JOR	Jordan	ZAF	South Africa
JPN	Japan	ZAI	Zaire
KEN	Kenya	ZMB	Zambia
KOR	Korea, Republic of	ZWE	Zimbabwe

## **STI WORKING PAPERS**

**1996**

1. Embodied Technology Diffusion: An Empirical Analysis for 10 OECD Countries  
*George Papaconstantinou, Norihisa Sakurai and Andrew Wyckoff*
2. The Impact of R&D and Technology Diffusion on Productivity Growth: Evidence for 10 OECD Countries in the 1970s and 1980s  
*Norihisa Sakurai, Evangelos Ioannidis and George Papaconstantinou*
3. Short-term Indicators: Using Qualitative Indicators to Update Production Indices  
*Paul Schreyer and Corinne Emery*
4. SMEs and Employment Creation: Overview of Selected Quantitative Studies in OECD Member Countries  
*Paul Schreyer*
5. Globalisation and Competitiveness: Relevant Indicators  
*Thomas Hatzichronoglou*
6. Factors Influencing the Steel Work Force: 1990 to 1995  
*Donald F. Barnett*
- 7 Measuring R&D in the Services  
*Alison Young*
- 8 The Evolution of Skills in OECD Countries and the Role of Technology  
*A. Colecchia and G. Papaconstantinou*

**1997**

1. Bibliometric Indicators and Analysis of Research Systems: Methods and Examples  
*Yoshiko Okubo*