# Unit-12:Growth and Obsolescence of Literature

## 12.0 : Objectives

After reading this Unit, you will be able to:

- explain and define the concepts of Growth' and Obsolescence' of literature;
- understand different kinds of growth models and patterns of obsolescence;
- use the measures for rate of growth and obsolescence of literature; and
- understand the implications of the growth and obsolescence of literature in library and information activities.

## 12.1 : Introduction

Dynamism is the dominant characteristic of science. The industrial and information revolutions have pushed and fuelled the growth of science. Since the mid-seventeenth century, it has grown at an ever-increasing pace from 'Little Science to Big Science' as referred to by. Derek De Solla Price. Researches in the sociology of science suggest several aspects of such growth anc development in scientific knowledge.

One feature which seems to be at the most centrality to the growth of science i; the rapid exponential increase of its literary output, other than those aspects such as, enhanced manpower, per capita investment in science, military o nuclear power, emergence of new fields, new departments, new institution; and technological advances.

It is this phenomenon of constant progress in the production of literature it science that makes it meaningful and useful to pursue the history of science Price pointed out that "what we must do in the humanistic and scientific analyze of science is search out such constancies of scientific method, of public reaction of the use of mathematical models and apply them to our criticism an, understanding of this science that seems so essentially modern...".

Understanding the growth of knowledge and its modelling has challenge information scientists (bibliometricians) and sociologists. This phenomenon can be measured qualitatively and quantitatively. Descriptive models of growth of knowledge use social phenomenon, which is quantitative in nature. Recentl3 the growth of knowledge has been measured using diffusion models a bibliometric tools of simple to complex statistical ones. The empirical statistics evidence has been drawn from many numerical indicators of various field and aspects of science. This has established that the normal mode of growth c science is exponential. That is to say that science grows at compound interest multiplied by some fixed amount in equal periods of time.

## 12.4 : Growth vs Obsolescence of Literature

Growth and ageing are two topics, that have received world wide attention and much reported upon. The influence of growth on the obsolescence of a specialty as expressed by the citation pattern has been the subject of many studies.

'Influence' of growth on obsolescence means that the obsolescence of literature is dependent on the time period under study; hence the obsolescence of a subject in itself is an undefined one. It is stated that 'obsolescence' is a function of growth. Experimental evidence is that faster the literature grows, the faster ii becomes obsolete.

There is no formal model for such an influence. Egghe has attempted to develop such a model. His work reports different findings for the synchronous and diachronous studies. Synchronous studies report that an increase of growth implies an increase of obsolescence, while the diachronous studies report al opposite pattern.

Such a study is not simple. Therefore, a theoretical and practical investigation of the influence of growth on obsolescence of literature is essential.

## 12.5 : Summary

This Unit discussed the

- concepts and definitions of growth and obsolescence of literature;
- models and measures of growth of literature;
- models and measures of obsolescence of literature;
- influence of growth on obsolescence of literature;
- implications of studies of growth and obsolescence;

To sum up, any study on the rate of obsolescence needs to specify user group, specify the literature (knowledge base), specify the sample size (the sampling difficulties in such attempts are formidable) and specify the measure/method.

# Unit-13:Science Indicators

## 13.0 : Objectives

In this Unit, you will learn about science indicators. When you complete it, you will be able to answer questions such as: Why do we need Science and Technology (S&T) indicators? How are they different from Research and Development (R&D) statistics? How did these indicators evolve? What are the functions and uses of different types of indicators? What are the sources of information for building S&T indicators? Why are output indicators considered more meaningful than input indicators? What are the problems one faces with literature-based indicators? How are patents data used in indicators useful in the corporate context? Do we need to develop S&T indicators for developing countries distinct from those used by advanced countries? How can you present complex indicator data in easy-to understand visuals?

## 13.1 : Introduction

> *"If you can measure that of which you speak, and can express it by a number you know something of your subject; but if you cannot measure it, your knowledge is meager and unsatisfactory*
>
> *- Lord Kelvin*

We are living in the knowledge millennium. The importance of knowledge, in particular scientific knowledge, is recognized like never before in the history of man. "Few would dispute the claim that a nation's science and technology base is a critical element of its economic strength, political structure, and cultural validity," says Eugene Garfield. Today, there is a tremendous desire in virtually every country, barring perhaps the absolutely poor countries, to step up investment in scientific and technological research and to reap the benefits of such research. However, the resources available, even in advanced countries, are rather limited and one has to make choices of areas to invest in. Investment on scientific and technological research and higher education in science, technology and medicine has to compete with other far more basic needs such as health, agriculture, primary and secondary education, defense, energy, and civic services. Furthermore, how does one allocate scarce resources, both human and financial, to different areas of research so as to maximize the returns, both social and economic? To improve the basis upon which these allocations are made, sound' methods are needed to assess the effects of scientific and technological activities on the country's economic and social development. Such assessment needs the development of a variety of indicators or measures of different aspects of the S&T enterprise. In the words of Hiroko Morita-Lou, indicators of technological development are needed for two main purposes: (i) to describe and facilitate the analysis of specific historical situations, and hence the elaboration of diagnosis, and (ii) to help identify inter-relations between different variables and phenomena where scientific knowledge is applied through technology; and to be used in policy formulation and planning.

## 13.2 : Indicators

The concept of indicators in itself is an interesting field. Some thinks are easy to measure directly, such as the mass or the speed of a cricket ball. Some phenomena - such as, quality of life, happiness, and innovation - defy direct measurement. Indicators are tools used to assist in the assessment of such phenomena. Often, assessments or evaluations make use of sets of indicators each of which may display or highlight some facet of the object of the evaluation. Such assessments and evaluations are important tools used by policy makers.

Almost always, statistics about the phenomenon under study provide the building blocks from which indicators are constructed. Often the two terms, statistics and indicators, are used interchangeably. As the literal meaning of the word indicates, indicators are indications of other things, usually each indicator emphasizing a specific aspect as well. As such, indicators are imperfect measurements, and therefore, several of them should be studied to discern trends. By examining a set of indicators one can discern trends in the behavior of the phenomenon at hand. Monitoring the fluctuation and movements of a number of indicators over time may reveal relationships between different elements or aspects of the phenomenon.

## 13.3 : Towards Science Indicators

Indicators for science and technology may be defined as "statistics which measure quantifiable aspects of the creation, dissemination and application of science and technology. These indicators should help to describe the science and technology system, enabling better understanding of its structure, of the impact of policies and programmes on it, and of the impact of science and technology on society and the economy." [Martin B. Wilk, *Stalislics Canatla, A Framework for Measuring Research and Development Expenditures in Canada, Catalogue 88-506E, preface.]*

There are many ways of looking at science. According to the *Concise DVf rrd English Dictionary,* science is the pursuit of systematic and ordered knowledge. It is a very complicated system of knowledge production and knowledge exchange. It can be viewed as a means of constructing models of reality. Science is a major human activity - a part of human culture in the contemporary world.

It has far reaching applications and implications in war and peace, in healthcare and agriculture, in transport and communication, in industry, in international trade and competitiveness - in short, in every aspect of human life. But, how does one measure such an activity as science? How does one measure knowledge or its impact on human life? These are indeed formidable challenges. If we cannot measure knowledge or science accurately, the next best thing to do is to measure all that is measurable about them. That is precisely what has happened. For example, we can pleasure the volume of research carried out in a country by the number of research papers published by scientists in that country, a surrogate for knowledge. We know that not all research papers are equal in merit, quality or utility. The two-page paper by James Watson and Francis Crick in *Nature* 1953 announcing the double helical nature of the DNA molecule or Einstein's 1905 paper on the special theory of relativity are far more profound than an average paper appearing even in the best of journals. And yet, for want of a better yardstick we settle for counting papers, and it is perfectly acceptable as long as the sample sizes are large enough to be statistically valid. Similarly, we can measure the amount of funds invested by a country on R&D and the number of scientists employed therein in R&D.

S&T data should not be considered in isolation. They need to be adjusted for the size of the economy and labour they represent. That is why some indicators are expressed as ratios of two entities a specific numerator over a general denothinator. For example, Gross Expenditure on Research and Development (GERD) over Gross Domestic Product (GDP). This indicator suggests that GERD and GDP are related non-linearly and that GERD/ GDP for smaller economies like Sri Lanka and Ghana is not to be expected to be close to the larger economies such as the USA and Germany.

In this Unit we will look at some of the important indicators that can give us an idea of the volume of research, the manpower involved, the financial allocations, international distribution and globalization of R&D, extent of collaboration, the level of innovation, etc. However, our thrust will be on literature-based indicators, which have played a major role worldwide.

## 13.4 : Historical Aspects

Since the early 1960s, there have been some concerted efforts at developing systematically indicators for measuring scientific and technological activities and their impact on society. The movement began in the United States, where concerned by the apparent decline in the relative performance of the country in science and technology, political leaders, economists and policymakers, skeptical of subjective judgements and anecdotal evidence, insisted on quantitative evidence. The idea was that similar to social, economic and financial indicators, science indicators would help one to appreciate the state of health of the scientific enterprise and be useful in charting out future programmes. By then economic indicators such as Gross Domestic Product (GDP), Industrial Production, Retail Sales, Unemployment Rates, and Consumer Prices, and financial indicators such as Money Supply, Interest Rates, Trade Balances, Exchange Rates and Foreign Reserves were well established and used by the major governments, banks, corporations, economists and analysts. The science indicators movement was preceded by sporadic attempts at constructing R&D statistics. John D Bernal and Stephen Dedijer were among the earliest in Europe to recognise the importance of building R&D statistics and quantitative indicators for the scientific enterprise. In June 1963, the OECD (Organization for Economic Cooperation and Development), a group of rich countries, held a meeting of experts on research and development statistics in Frascati, Italy. The deliberations resulted in the first official version of the "Proposed Standard Practice for Survey of Research and Development", better known as the *Frascati Manual,* Around the same time Unesco had arso started collecting comparative statistics on R&D efforts in a number of countries. The fifth edition of the *Frascati Manual* was published by OECD in 1993. The *Manual is* one of the cornerstones of OECD's effort to increase understanding of national science and technology enterprises and systems of innovation. However, it deals exclusively with what are known as **input indicators -** human and financial resources devoted to R&D. Over the years these indicators have been found very useful. For example, the OECD *Reports on Science and Technology Indicators* and OECD's *Science and Technology Policy Review and Outlook* series provide useful measures of the scale and direction of R&D in many countries, sectors, industries, scientific fields, etc. To these input indicators were added output indicators, such as papers published and patents applied/granted, and impact ofresearch on other research, economy and society, and esteem indicators, such as awards and medals won, election to Fellowships of Academies, appointments to named chairs, and so on. The works of Fritz Machlup (Production and distribution of knowledge in the United States, 1962) are also noteworthy.

Today many countries and international organizations produce science and technology indicators. Notable among them are the *Science and Engineering Indicators,* published biennially since 1972 by the National Science Foundation of the USA, and the *European* ***Report on*** *S&T Indicators,* published since 1974 by the European Commission. Several European countries, such as France and the Netherlands, have set up Observatories of Science and Technology. In India, the Department of Science and Technology has a division called the National Science and Technology Management Information System, ,which produces valuable statistical information on science and technology in India. Australia, New Zealand, Canada, the Nordic countries, the OECD countries, Israel, and many Asian countries have been using S&T indicators to different extents as a measure of their scientific and technological development and simultaneously measure their competitiveness. All these agencies / countries attempt to develop a comprehensive system of indicators for describing the state of the entire scientific endeavour and to chart its changing state.

The science indicators movement was greatly helped by the publication in the early 1960s of the *Science* ***Citation Index*** by the Institute for Scientific Information, Philadelphia, PA, USA, founded by the well-known information scientist Eugene Garfield. *SCI* forms the major source of literature-based indicators today. The multidisciplinary database, PASCAL, published by INIST, France, is also used, at least in Europe, for construction of indicators. Since the early 1980s CHI Research, Inc., an American company founded by F'ranccis Narin, is active in the business of constructing literature-based indicators. This company uses *S'17* data as well as US patents data to construct several indicators for the US National Science Foundation and many other clients - both government and corporations. Many academic institutions, such as the Information and Scientometrics Research Unit of the Hungarian Academy of Sciences, Budapest, Science Policy Research Unit at the University of Sussex and the Centre for the Study of Science and Technology at Leiden University, have made substantial contributions to this field.

## 13.5 : Functions of Science Indicators

What are science indicators and how-are they different from R&D statistics? Let us see what the experts say.

Science and technology indicators, says a 1979 OECD document, are a series of data designed to answer questions about the current state of and/or changes in the science and technology enterprise, its internal structures, its relationship with the outside world, and the degree to which it is meeting the goals set for it by those within or without.

In the words of Giorgio Sirilli of Italy's National Council of Research, "The purpose and function of science indicators are to follow changes in the scientific enterprise and its components over time, and thereby to reveal strengths and weaknesses as they begin to develop. Such indicators, updated regularly, can provide early warnings of trends and, taken together, can make decision-makers more aware of the interrelationships of the many variables that describe a scientific system's effort. Hence they can assist those who set priorities for the enterprise and allocate resources to it". [See Hiroko Morita-Lou (ed), 1985. *Science and Technology Indicators for Development*, UNCSTD and Westview Press, Boulder, p. 188].

"Indicators are not simply statistics, and statistics are not ipso facto indicators -unless some theory or assumption makes them so by relating the indicator variable to a phenomenon that is not what it directly and frilly measures" [H Morita-Lou, p36].

According to Gerald Holton, the Harvard University physicist, the term "indicator" should be reserved for a measure that explicitly tests some assumption, hypothesis, or theory; for mere data, these underlying assumptions, hypotheses, or theories visually remain implicit. Indicators are the more sophisticated result of a complex interaction between theory and measurement. The difference between data and indicators is best illustrated by an analogy. The measurement of the rapidly declining blood pressure of a patient remains as data until it is seen by someone with enough understanding of physiology to recognise it as an indicator of a change in the state of health. [Holton in Elkana *et al.*, 1978].

## 13.6 : The Use of Science and Technology Indicators

The primary purpose of S&T indicators, which are often used in conjunction with other econo.nic indicators, is to guide the decision of S&T policy makers. Systematic and appropriate S&T Indicators would provide the policy makers with an indication of the status of the national S&T system and the relationship between scientific and technological efforts and economic growth. They can serve as instruments of planning, evaluation, and resource allocation.

S&T indicators that explain historical trends or relationships between variables at a point in time may be useful in determining the nature and level of resources to be allocated to scientific and technological endeavours and to sectors that support such activity. They may also provide the necessary data for assessing past performance in capacity building and utilization.

With the increased importance of S&T in a nation `s economic system, S&T Indicators have been developed to describe and map the national system of innovation so that national competitive advantages could be identified and promoted.

### 13.6.1 : S&T Indicators for the Developing Countries

We live in a divided world. A few rich countries form the centre and most of the rest form the periphery. The OECD countries and the European Union constitute the centre and the developing countries of Africa, Asia and Latin America constitute the periphery. Right from the beginning, science indicators have been developed and used mainly by the advanced countries of the world, almost all of which have been benefited by modern science and technology. For the developing countries, whose concerns are different from those of the economically richer countries, it is necessary to formulate a fairly complete system of socio-economic development objectives as well as a fairly complete system of indicators. The latter should be able to reflect the needs and requirements of the objectives so as to ensure their successful realization.

Indicators to be developed should be appropriate to the characteristics of the local social system, whether this is the national economy or the S&T system. The different ways similar types of activity are organized in different settings means that identical statistics are not necessarily indicators of the same or similar things. S&T indicators that prove useful in the USA and the European Union may be misleading and irrelevant in Sub-Saharan Africa. Indicator development must be related to research on the concrete characteristics of the formal and informal systems of S&T in the types of society concerned.

Ideally, in the context of developing countries, one would expect S&T indicators to:

"provide insights on the impact of S&T efforts on national socio-economic development;
help countries inventory to assess their S&T capabilities;
identify the extent to which a country's S&T system is integrated into or dependent upon the world system;
help set priorities between different disciplines, institutions, or research groups competing for scarce funds, based on an assessment of their recent S&T performance; and
help research institutes better manage their research programmes."

[Hiroko Morita-Lou, p25]

The formulation of indicators for given countries or part thereof, deserves a higher priority than of indicators that facilitate cross-country comparisons. This is simply because review of planning and policy formulation is more meaningful and operational in terms of national action.

A comprehensive indicator system for developing countries should include:

not only economic indicators, but also social, S&T, environmental, and quality of life indicators;
not only these indicators *per se*, but also relationships among them, especially the impact of S&T on socio-economic developments;
not only indicators of quantitative measurement, but also qualitative indicators of socio-economic benefits.

In India, for example, the Centre for Science and Environment, is spearheading the development of environment indicators appropriate to the country's needs, and the M. S. Swaminathan Research Foundation (in Chennai) is looking at how best to adapt modern technologies to reach the unreached.

## 13.7 : Types of Indicators

Indicators used in the assessment of S&T can be of different types, and so it is necessary to make some *a priori* choices as to which types one wants to utilize in one's evaluation scheme. The following are the elegant topology of types of indicators.

### 13.7.1 : Input Vs Output

Input-output analysis came to prominence with the work of the economist Vassi ly Leontief. It is much easier to measure input than output. For example, one can measure the amount of funds invested in research, the number and kinds of institutions established for scientific activities, and number of scientists trained and employed for research, etc. On the other hand, the output of scientific knowledge is rather intangible and hence hard to measure. In case of technology, its product may be somewhat more concrete, but gauging its quality is also difficult. S&T outputs indicators attempt to relate investments in S&T and their possible benefit to the country. Even though there could be other outputs from investments in R&D and human resources, statistics on patents applications/granted and data on bibliometrics (number of S&T publications produced and cited) are the two most used S&T output indicators. Patents would measure the level of innovation and inventiveness while bibliometric indicators assess the level of S&T of a country as well as providing an indication of the quality and capability of local scientists.

Even in describing input, however, some difficulties arise. For instance, although it is easy to give a simple headcount of the number of scientists employed, it is much more difficult to describe the quality distribution within such scientific manpower. Moreover the classical input-output analysis has some limitations, including the difficulties of measuring proximate output and of relating input to output. The output of one system may be the input of another. Human resources' in R&D is the output of S&T education but the input to R&D activities. Publications are the output of R&D but the input to technological development work.

Sometimes input indicators are implicitly used to measure the output. We know, however, that the above assumption is a very misleading one to make. For a functional and realistic evaluation of the S&T system, output indicators must be measured directly and not through input indicators.

S&T indicators are no doubt more meaningful if they are related to outputs rather than inputs. Input indicators say little about the contribution of science and technology to economic aggregates like GNP and its growth over time. For example, the number of scientists and engineers involved in R&D may give some hints for the planning requirements of education, S&T, and R&D institutions. But it does not give any qualitative or efficiency indication of how scientific manpower inputs contribute to meeting objectives such as the fulfillment of basic needs and employment generation. Likewise, the number of doctors in a country tells us nothing about the geographical distribution of these doctors, their specialization, or supporting medical services. Yet this is what would be required if we were to measure the contribution of medical S&T to the satisfaction of health needs of given target population

### 13.7.2 : Quantitative Vs Qualitative

Some indicators can be strictly quantitative, and we can assign a definite value on some definite scale to describe some aspect of the S&T activity. For example, the number of scientists working in a given institution or on a given project. Other indicators may be predominantly qualitative. For example, through peer review, we may come to the conclusion that the scientists working in a laboratory are (or are not) good enough for carrying out a project.

Quantitative indicators have a definite intellectual appeal, giving the impression of being more objective, more precise, and more reliable than mere qualitative impressions. Indeed, one should aim at developing quantitative indicators for as many aspects of S&T as possible. A quantitative indicator is no better than the reliability of the set of assumptions and models on which it is based, and no better than the reliability of the data used in arriving at the quantitative mark. Some very crucial elements in the S&T complex may not be amenable to quantification. Indeed, some sociologists claim that, at least at the moment, only the most superficial and trivial aspects of S&T can be quantified, and hence restricting oneself to quantitative indicators alone amounts to disregarding some of the most essential aspects of the overall problem. It is therefore advisable to use both quantitative And qualitative indicators, wherever possible.

### 13.7.3 : Activity, Productivity and Progress

In describing what we want to find indicators for, it is useful to distinguish among the three concepts of scientific (or technological) activity, productivity, and progress. Productivity would mean going a step further from activity and assessing whether all the activities involved in a particular setting actually advance the attainment of an eventual goal or not. That is, we would have to ascertain that the activity carried out is in the right direction and is done well enough to move towards the given goal. Finally, progress would consist of measuring the degree to which all the productivity described brought us closer to our task. It is clear from this picture that while what would matter most is the measurement of progress, what is easiest to do is the measurement of activity.

### 13.7.4 : Quality, Importance and Impact

Another useful three-way distinction can be made among indicators along the trichotomy: quality, importance, and impact. The quality of research refers to how well it has been done, whether it is elegant, error free and general, and how intriguing the results are. It is a relative indicator and obviously different people may judge it differently. Importance denotes the potential influence on related research activities under ideal conditions, extrapolated into the unknown future. Impact means the actual influence that piece of research will have on the course of science and its various aims and use in the future, something that may be influenced by many factors, many extraneous to science and its various aims and applications. Although our ultimate goal would be to find indicators for impact, often we may have to be content with having indicators only for importance or only for quality.

### 13.7.5 : Functional Vs Instrumental Indicators

We can develop indicators for certain functions of S&T activities (e.g., for research results), or we may want to evolve indicators for certain tools and instruments that are used in the S&T infrastructure (such as for the upkeep and maintenance of scientific equipment or for management methods of scientific research). The output-input dichotomy has nothing to do with the functional vs instrumental dichotomy. While in the ultimate analysis what matters more are the functional aspects of the S&T system, for diagnostic purposes instrumental indicators may be very useful and in fact even superior to purely functional ones in that they may more readily point at various remedial measures that can be taken to eliminate certain deficiencies.

## 13.8 : Validity and Reliability of Indicators

It should be possible to check the validity and the reliability of indicators. Often, indicators are obtained from compilations of statistics, some of which are not reliable due to the method by which the data were collected, to implicit assumptions that may not be appropriate, or to other reasons. The validity and reliability of an indicator may be gauged by checking for consistency with other indicators. For example, it is known that there are a number of relationships among manpower indicators and financial indicators. If a set of indicators shows large deviations from the expected relationship, validity and reliability may come into question. This calls then for deeper probing into the reasons for such a deviation, including possible errors in methodology or possibly dubious assumptions made in the course of data collection. Further analysis may not always yield a concrete answer, but at least a certain degree of critical assessment of the data is provided.

## 13.9 : Building S&T Indicators

Before we can build S&T indicators, it is necessary that we understand the way science and technology are organised and administered. Scientific research and technological innovations are carried out by qualified (and often university-trained) researchers in organisations such as government-owned (or funded) or corporate laboratories, universities and non-governmental organisations around the world. Research is expensive and very often researchers have to obtain research grants from funding agencies, sometimes from outside their home countries. Their own institutions cannot afford to absorb the entire cost. Different researchers and their institutions choose different areas to work. Some are inclined towards theory and others may work in experimental sciences. Some may work in basic or fundamental research and others may engage in applied or developmental research. There are occasions when scientists find the need to collaborate with others, especially where they feel the skill sets of the collaborators will complement their own and facilitate solving a problem. In essence, the scientific enterprise has three components: people, funding, and what is being done. We will learn more about this aspect of science in the Unit on sociology of science and scientometrics. It is these aspects that S&T indicators aim to measure.

### 3.9.1 : Definitions

We will begin our study of S&T indicators with a few definitions from the *Frascati Manual* and other documents in the section on Further Reading.

1) Research.and experimental development (R&D) comprise creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications. R&D covers three activities: basic research, applied research and experimental development.

2) Scientific and technological activities comprise "...systematic activities which are closely concerned with the generation, advancement, dissemination and application of scientific and technical knowledge in all fields of science and technology. These include such activities as R&D, scientific and technical education and training (STET) and the scientific and technological services (STS)..." (Frascati Manual, p. 18). ,

3) The STS are defined as "...activities concerned with research and experimental development and contributing to the generation, dissemination and application of scientific and technical knowledge".

4) The term "R & D statistics" covers a wide range of possible statistical series measuring the resources devoted to stages of R & D activity and the results of that activity (Frascati Manual, p. 18).

5) "Basic research is experimental or theoretical work undertaken primaril: to acquire new knowledge of the underlying foundations of phenomen and observable facts, without any particular application or use in view' (Frascati Manual, p. 68).

6) "Applied research is also original investigation undertaken in order tt acquire new knowledge. It is, however, directed primarily towards a specifn practical aim or objective" (Frascati Manual, p.69).

7) "Experimental development is systematic work, drawing on existing knowledge gained from research and practical experience, that is directec to producing new materials, products and devices; to installing nevi processes, systems and services; or to improving substantially those alread) produced or installed" (Frascati Manual, p. 70).

8) Gross Domestic Product (GDP) is the total flow of goods and service; produced by an economy over a specified period of time, normally one year.

9) Gross National Product (GNP) is the sum of GDP and the income accruing to domestic residents arising from investment abroad less income earned in the domestic market accruing to foreigners abroad.

10) Scientific and technological innovation: Scientific and technological innovation may be considered as the transformation of an idea into a new or improved product introduced on the market or a new improved operational process used in industry and commerce or into a new approach to a social service (Frascati Manual 20).

## 13.10 : Literature based Indicators

We will now move on to different types of indicators, beginning with literature-based indicators. These are categorized under three heads: Publication indicators, Citation indicators and Patent indicators.

### 13.10.1 : Publication Indicators

1) **Publication counts** indicate the number of publications in which the given country or a country of the given region occurred at least once in the corporate address field. Publication types other than article, letter, note and review are often omitted. This indicator reflects scientific output. How do we assign multi-author papers? Some

assign such a paper to each institution / country whose address is listed. The same paper may be assigned to more than one institution / country and therefore the total number of papers will be larger than the real number. Others follow the fractional count method. The total number will tally with the actual number, but this procedure is somewhat cumbersome.

2) **Science Production Index** (SPI) defines the percentage share of world output: the total number of articles published in a given country, in a given period of time and in a specific field of science, divided by the total number of articles published in the world (WRD), in the same field, over the same period of time.

$$SPI = \frac{\text{Number of articles of country A}}{\text{Number of articles of the WRD}} \times 100$$

3) **Activity Index (AI)** :

$$AI = \frac{\text{the world share of the given country (region) in publications in the given field}}{\text{the overall world share of the given country (region) in publications}}$$

4) **Relative Specialisation Index (RS1):** The Relative Specialisation Index indicates whether a country (or region) has a relatively higher or lower share in world publications in a particular field of science than its overall share in world total publications. RSI is closely related to the so-called Activity Index (AI) long used in bibliometrics, which, in turn, is a version of the economists' "comparative advantage index."

$$RSI = \frac{AI - I}{AI + I}$$

## 13.10.2 : International Collaboration Indicators

1. COA (Coauthorship) is defined as international institutional coauthored articles. This output indicator measures the number of such articles resulting from cooperation. It is a measure of productivity. For example, the share of internationally coauthored papers in the UK rose from 14.0 % in 1985 through 18.8% in 1990 to 26.2% in 1995. During the same period, India was involved in much less international collaboration: 8.3% of papers in 1985, 10.1% in 1990 and 14.7 % in 1995.

   One can also count papers coauthored by researchers from different institutions within the same country (domestic collaboration). For example, in the three years 1992, 1995 and 1998, about 10% of Israel's publication output in new biology had resulted from domestic collaboration, as against 42% from international collaboration.

2. **COP** (Cooperative effort) measures the cooperative efforts made by countries through writing articles and the **number of scientific links** created between countries through common work. Thus COP measures the degree of the scientific ties between countries and the degree of participation of a country to the international scientific community. If there is a paper from the Raman Research Institute in Bangalore with coauthors from USA, Japan and Germany, then it will be counted as one paper for India and three links - one each with USA, Japan and Germany. In short, if there are n countries in the byline of a paper, there are n-l links.

   As each country has a right to count its participation in an internationally coauthored paper, the total number of links will *ben x (n-1),* where *n* is the number of countries represented in each of a country's coauthored papers. As seen from *SGI* 1998, India had published 11,437 papers. Of these 2,015 were written in collaboration with authors outside the country. But, there were 3,300 links, meaning in many papers there were authors from more than two countries.

3. **International Cooperation Index,** also known as **Salton's measure,** is defined as the number of joint publications (involving A and B) divided by the square root of the product of the number of total publications (of A and B), where i and k are two large publishing entities, such as two countries or two institutions.

$r_{AB} \doteq n_{AB} / (n_A n_B)^{1/2}$ ,

where $r_{AB}$ is the cooperation index between countries A and B, and $n_A$ and $n_o$ are the number of papers published by the two countries and $n_{AB}$ is the number of papers coauthored by scientists from both countries. In 1998, India (11437 papers) and China (13878 papers) had collaborated in only 71 papers. That would mean an international cooperation index of

71/(11437 x 13878)$^{1''2}$ = 0.005636

In contrast, the UK (57349 papers) and Germany (54676 papers) had collaborated in 2531 papers. Hence the international cooperation index between the two countries is:

2531/ (57349 x 54676) $^{''2}$ = 0.0452

4. **Affinity Index** (AFI) is a measure of the amount of collaboration between a given country (A) and another (B), compared to the total collaboration of the given country (A) with the entire world, in a given field of science, during a given period of time. AFI is therefore: the number of COPs between A and B divided by the total COPs A has with the rest of the World (WRD), in a given field and during a given period of time. It indicates the scientific affinity of A towards B (A-- .B).

India and Canada have collaborated **in 88** papers in all of science, as seen from SCI 1998. India has collaborated in a total of 2015 papers in the same period. The affinity India has for Canada is then

88/2015 = 0. 04367

Often AFI is multiplied by 100 for convenient representation.

5. **Internationalisation Index** (INI) is a percentage of total number of COA of a given country published during a given period of time in a given field over the total number of articles published in the same given country during the same period of time, in the same field.

In all of science, India had collaborated in 2015 papers out of its total of 11437 papers, as seen from SCI 1998. Hence India's internationalisation index is:

2015/ 11437 = 0.1761.

## 13.10.3 : Citation-based Indicators

Citation counts may be considered as 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. One can also use the ratio of citation count to publication count as an indicator of citation rate per publication. Citation data are usually obtained from one or the other citation index products of the Institute for Scientific Information. When dealing with large volumes of publications, say the output of a country,

one can think of three different citation measures, viz. observed citation rates, expected citation rates and relative citation rates.

1) **Observed citation rate** is the actual number of times a paper is cited. The mean observed citation rate is the number of citations per publication.

2) **Expected citation rate** of a paper is the average of the number of citations of all papers published in the same journal in the same year, which in turn depends on the impact factor of the journal.

3) **Relative citation rate** is the ratio of observed citation rate divided by the expected citation rate. The relative citation rate is a measure of whether the publications of a country attract more or less citations than expected on the basis of the average citation rates of the journals in which they appeared.

In a series of papers in *Scientometrics*, Tibor Braun and colleagues at the Information and Scientometrics Research Unit of the Hungarian Academy of Sciences have calculated the observed, expected and relative citation rates for more than seventy countries and hundreds of journals.

## 13.10.4 : Problems Associated with Collection of Literature

1. **Homonym Problem.** In looking through the *Science Citation Index* to identify papers citing the work of G. N. Ramachandran, the well-known molecular biophysicist, one must be careful not to select citations to papers by other people named G. N. Ramachandran. This problem can be overcome if one has the complete list of the concerned scientist's publications with full bibliographic details.

2. **First Author Problem.** The citation index databases of the Institute for Scientific Information, and *Mathsci* database of the American Mathematical Society provide addresses of all authors of papers they index. But most other secondary services only provide author information for the first author of a paper. One must be careful in using such databases not to overlook the existence of co-authors. Also if one is using *SCI* for collecting citations to the work of a scientist, one must look for all papers by the scientist, even those where he/she is not the first author. Many professors give first author status to their doctoral students and if one looks for citations to their work giving their names as the search terms one would end up with only a partial list of citations to their work.

3. **Journal Quality Problem.** There are tremendous variations in the quality of journals. If one wants to take these variations into account, one can weight individual papers published in a journal with, say, Garfield's impact factor (available in the print and CD-ROM editions of the *Journal Citation Reports)* or influence weight evolved by Pinski and Narin. [In calculating the impact factor ofa journal, all citations to papers published in the journal are treated as equal, whereas in calculating the influence weight one takes into account the impact factors of the citing journals as well. Even this is not adequate, as not all papers published in a journal are cited the same number of times.

4. **Citation Lag Problem.** There are substantial field-to-field variations in how rapidly papers get cited. Fast moving fields have short lags between the time a paper is published and then cited, while in slow moving fields, the lag is considerably longer. If one is unaware of these field-to-field differences, one may misinterpret one's finding. Added to this problem is the delay in journals published from non-mainstream countries and in non-English languages getting noticed in mainstream scientific countries performing much of world's scientific research. Often, these journals have a very small print run and are not subscribed to by major academic and research libraries as are mainstream journals published in the advanced countries. A vast majority of scientists may come to know of papers published in these journals only when they make a literature search using secondary services, which in turn cover these journals a few months after they appear in print.

5. **Institutional Address Problem.** Different abstracting and indexing services follow different policies in reporting the institutional addresses of authors, ranging from no institutional information to detailed information. The citation index databases of ISI give full addresses of all authors, whereas some of the better-known STM abstracting services fail to give the country name in a substantial percent of documents they index. This is particularly of concern to bibliometricists trying to characterise the output of a country. To overcome this problem, if one is working with a CD-ROM database, one should give not only the name of the country - say India - in the address field but also all the cities and towns Ludhiana, Kharagpur, Aduthurai - where possibly some research would have been carried out, in a proper Boolean form.

6. **Abstract/Index Coverage Problem.** Some abstracting and indexing services are very consistent in their journal and article coverage policies. The *SCI,* for example, covers all papers appearing in a well-defined set of journals. Other abstracting / indexing services may not cover all that is published in a journal; some may include only those pieces that happen fo come to their attention. Occasionally, journals branch off into sections, e.g. *Indian Journal* of *Technology* branched into two sections in the mid-1990s. There are also mergers of two journals into one. In such cases, computing impact factors of these (split and merged) journals would pose problems. Also, the

abstracting/indexing. service may cover a journal in a particular year but drop it the next year. For example, at one time or the other *SCI* has indexed 40 Indian journals, but in 1999 *SCl* covered only a dozen Indian journals. Some developers of indicators, notably Francis Narin and his company CHI Research Inc., have overcome this problem by defining a `constant set' of journals, i.e. those journals that are covered by the database in each one of the years under study.

7. **Problem of journal title abbreviations.** Often authors use non-standard abbreviations of journal titles when they quote earlier literature. Database producers such as ISI and Chemical Abstracts Service (who have started providing citation information in their web product *SC'IFINDER)* cannot obviously spend time and energy to correct such mistakes. Journal titles are often long and to save space and composing effort it is customary to use abbreviated titles of journals, e.g., *J. Sci. Ind. Res. for Journal of Scientific and Industrial Research.* Different secondary services render these abbreviations in different ways. For example, *Afathsci* (in its CD-ROM and web versions) renders them with a hyphen between every part, e.g., *Phys. -Rev. -Lett.* for *Physical Review Letters.* Other databases may not use hyphens between parts. These variations become critical when one has to process large volumes of data using a computer. Imagine one wants to study fisheries research in India and wants to make the study truly comprehensive. Fisheries research is covered by many databases - *Science Citation Index, Biochemistry and Biophysics Citation Index, Biosis, Aquatic Science and Fisheries Abstracts, CAB Abstracts, AGRICOLA*, etc. Although all of them give bibliographic information for articles they index, *not* all articles are indexed by all of them and not all of them follow the same style of giving the information. Each one of them has its own structure and pooling all the unique entries into a single database is truly a formidable task.

## 13.11 : Indicators Based on Patents

Scientists, inventors and corporations protect their inventions by taking patents in order to reap the financial benefits which stem from their invention or innovation. Patenting gives an invention a legal status and provides protection from unlicensed copying. Within companies, patenting may sometimes serve more strategic aims, for example by lending greater status and credibility to inventions and thus helping to attract further internal or external funding. At the same time, patenting leads to the publication of the technical details of the invention. Patent data are therefore publicly available, cover a wide range of technologies, and usually extend over a long time period. In being so closely linked to invention, and so easily accessible, data on patent applications are often considered as an a ctremely useful proxy for technological output.

The following ratios are widely used as patent indicators:

1) Patents per population
2) Patents per GDP
3) Patents (extended abroad) *per* value of export
4) Patents per total (or business enterprise) R&D expenditure
5) Patents per total (or industrial) researchers.

6) Domestic patents denote the number of patents taken out in a country by residents of that country. This is a measure of innovative activity within the country.
7) External patents denote the number of patents that residents of a country take out in other countries. This represents corporate assessments of a country's ability to use and market a technology.
8) Patent growth percent in a subject area is the change in the number of patents from one period to another, expressed as a percentage. This identifies technologies receiving increasing emphasis and those in which innovation is slackening off. It also identifies companies increasing/ decreasing their technological development.
9) Percent of company patents in a subject area is the percent share of patents in a technology area in the total number of patents for that company. This indicates which technologies form the core of a company's intellectual property portfolio.

### 13.11.1 : Patent Citations of Scientific Papers and Prior Pa

Patents can also give an idea of science/technology linkages. Citations of scientific papers in a patent application refer to prior art relevant to the patentability of the application. Although some questions have been raised about citing patterns for scientific papers, as they may be affected by the examiner's expertise, the nature of the particular field and a variety of other factors, for certain technology fields, especially the emerging ones, examiner's citations of scientific papers do seem to be relevant in tracing linkages between technology and science. More exactly, such citations' "weight", both quantitative and qualitative, does attest the extent of the scientific basis for any given technology. At least three indicators can be used to assess that extent:

1) **Numbers of scientific papers cited:** more the number of papers are cited, the closer the linkages between technology and science;

2) **Average lead time between publication of papers cited and patent grants:** the shorter the time gap, the closer the linkages between technology and science;

3) **Types of journal in which cited papers appear:** the more such journals publish fundamental research findings, the more fundamental and the scientific knowledge on which the patents depend.

CHI Research Inc. has developed many advanced patent based indicators by analyzing the references to earlier patents on the front pages of later patents, or "patent citations". These indicators are derived from the simple idea that references placed on patents to help establish the novelty of the invention represent "prior art". Research has established that highly cited patents represent economically and technically important inventions.

### 13.11.2 : Limitations of Patent based Indicators

OECD Patent Manual 1994 lists the following inadequacies of patent-based indicators:

- The propensity to patent varies across enterprises, industry sectors and countries.
- For certain firms in particular sectors, there may be other more effective means of appropriating the market benefits of their inventions, such as secrecy, rapid launching or low pricing.
- Not all inventions can be patented, for example computer software.
- The requirement of novelty for the granting of patents means that certain more incremental technological innovations will not be covered (which is important for all countries, but particularly so for less developed countries, whose technological activity may not be adequately portrayed by patents).
- For a complex invention, the number of patent applications which need to be filed may vary from one patent office to another.
- Each patent system - either national systems such as the US Patent and Trademark Office (USPTO) or regional systems such as the European Patent Office (EPO) - provides protection for a specific geographic area. Therefore, export and market strategies will strongly influence the specific patent office(s) at which a patent application is filed. Consequently, for each geographic area, there is always a bias towards domestic applications.

## 13.12 : R&D Expenditure

Expenditures on R&D may be spent within the statistical unit (intramural) or outside it. The basic procedures for measuring the expenditures are:

- to identify the intramural expenditure on R&D performed by each statistical unit;
- to identify the sources of funds for these intramural R&D expenditures.

Large amounts of money are Invested on R&D every year. Expenditure on R&D performed in the United States topped $200 billion in 1997. The investment came not only from the US government (i.e. tax payers), but also from the industry (corporations). A few indicators based on investment used in mainstream countries are given below:

## 13.12.1 : Gross Domestic Expenditure on R&D

GERD is the total intramural expenditure on R&D performed on the national territory during a given period *(Frascati Manual, p. 101).* It is constructed by adding together the intramural expenditures on R&D as reported by the performing sectors. The GERD and GERD matrix are fundamental to internal examination and international comparisons of R&D expenditures. To summarise, the GERD serves as a general indicator of S&T activity and not as a detailed inventory of R&D projects within an organization, sector, or country. It is an estimate and as such can show trends in R&D expenditures by sector and sub-sector, by region and country, from year to year. The GERD estimates are sufficiently reliable for their main use as an aggregate indicator for science policy.

## 13.12.2 : Gross National Expenditure on R&D

GNERD is an optional supplementary aggregate which comprises total expenditure on R&D financed by institutions of a country during a given period. It includes R&D performed abroad but financed by national institutions or residents; it excludes R&D performed within a country but funded from abroad. It is constructed by adding the domestically financed intramural expenditures of each performing sector and the R&D performed abroad but financed by domestic funding sectors *(Frascati Manual*, p. 101).

## 13.12.3 : Intramural Expenditure

These are all expenditures for R&D performed within a statistical unit or sector of the economy, whatever the source of funds. Expenditures made outside the statistical unit or sector but in support of intramural R&D are included. Both current expenditure such as employment costs or expenditure on materials and capital expenditure, for example on buildings or equipment, are included. The sum of the intramural expenditure on R&D across all reporting units or sectors is equivalent to gross domestic expenditure.

## 13.12.4 : Government R&D Appropriations

GRDA denotes all appropriations by central government allocated to R&D in central government budgets. Data on government R&D appropriations therefore refer to budget provisions, not to actual expenditure. The figures on actual expenditure, which are not available in their final form until some time after the end of the budget year concerned, may well differ from the original budget provisions.

## 13.12.5 : R&D Intensity

R & D intensity applied to corporations is a measure of the extent' to which sales revenues are reinvested in R&D.

## 13.13 : Human Resource Indicators

The data on human resources indicators are useful for examining the stocks and the flows of highly skilled personnel particularly scientists, engineers and technicians. It is now increasingly accepted that the definition of human resources indicators should also include post-secondary teachers, medical practitioners, consultants and R&D managers. They cover both categories namely qualified personnel who are not employed in S&T related activities and those without formal qualifications but involved actively in skilled work.

The following definitions are adopted Researchers (or scientists and engineers) are professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems, and the management of the projects concerned. Technicians and equivalent staff are persons whose main tasks require technical knowledge and experience in one or more fields of engineering, physical and life sciences, or social sciences and humanities. Technicians perform their tasks under the supervision of researchers.

Initial coverage of R&D personnel: All persons employed directly on R&D should be counted, as well as those providing direct services such as R&D managers, administrators, and clerical staff. Some people may not be engaged in research full time. For this reason, 'full time equivalent' is defined as if two persons devote half their working time to research, then they are counted as one full time equivalent.

In some countries, every one with a degree (or diploma) in science, engineering, technology, medicine or

agriculture, irrespective of whether the person is actually engaged in some scientific or technical activity, is included in the number of S&T personnel. That is how India is considered to be having the third (or second) largest S&T manpower in the world.

R&D personnel can be classified by level of education (B.Sc, M.Sc, Ph.D, B.Tech, M.Tech, etc.), by field of expertise, by nature of work, age, sex, etc. One can also develop indicators for migration of qualified scientists and engineers. For example, thousands of Asian computer professionals leave their home countries every year to work in North America and Europe.

## 13.14 : Esteem Indicators

Scientists work not just for money. They work for recognition. They want other scientists to read and cite what they write. They are happy when they get elected to Academies such as the Indian Academy of Sciences (Bangalore), the Royal Society (London) the National Academy of Sciences of the USA (Washington DC) or the Third World Academy of Sciences (Trieste) or appointed to named chairs in major universities, They work hard to win awards such as the Bhatnagar Prize or the Nobel Prize. Different fields have their own awards and medals. Mathematicians aim at the Field Medal. Researchers in the field of scientometrics would like to win the Derek de Solla Price Medal. There are many awards and other kinds of recognitions to recognize achievements at different levels. These recognitions can form the basis of indicators. For instance, at one time there were five Bhatnagar Prize winners at the Institute of Mathematical Sciences, Chennai, where the total faculty strength is less than 100. More than a dozen people associated with Lord Rutherford went on to win the Nobel Prize. Professor M.S. Swaminathan was rated by Time Magazine as one of twenty Asians of the Twentieth Century.

## 13.15 : Visual Presentation of Indicators

Many of the indicators, especially when one uses them to compare different countries/ institutions, are presented as diagrams and charts. Given here are four figures to illustrate how visualization is used in the field of S&T indicators. Many more can be seen in the references listed under the section `Further Reading.'

Fig. 1 An example of the use of pie chart to depict the distribution of articles in world scientific and technical journals, by region/ country [Based on data *from Science Citation Index* 1995]. Source: National Science Board -*Science and Engineering Indicators -1998*, National Science Foundation, Arlington, VA, USA, 1998 (NSB 98-1), p. 5-

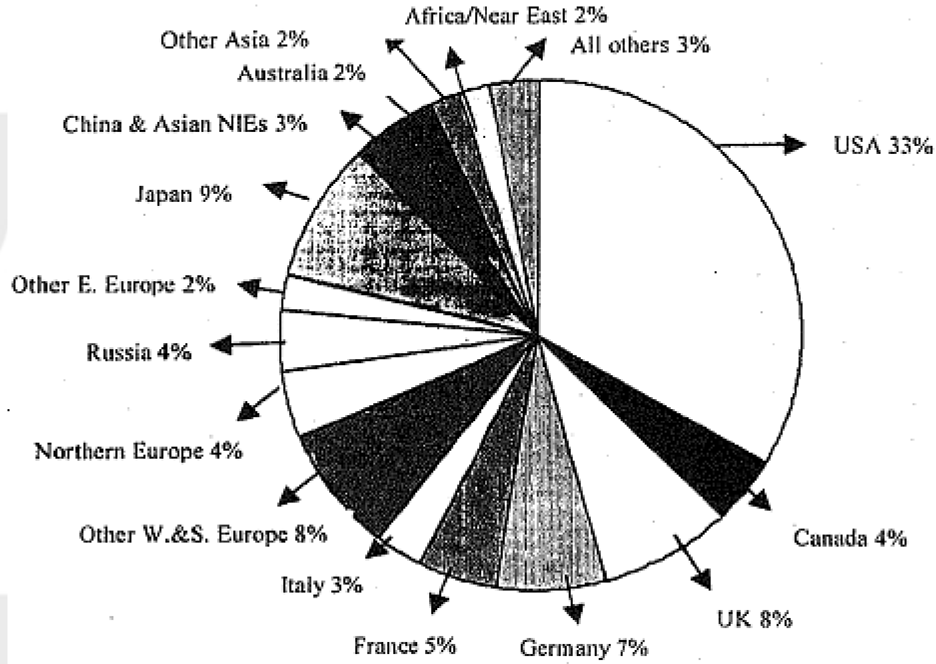



Fig.2 An example of the use of column chart to depict the distribution of articles in world scientific and technical journals, by region/ country (Based or the Data at Fig. 1)

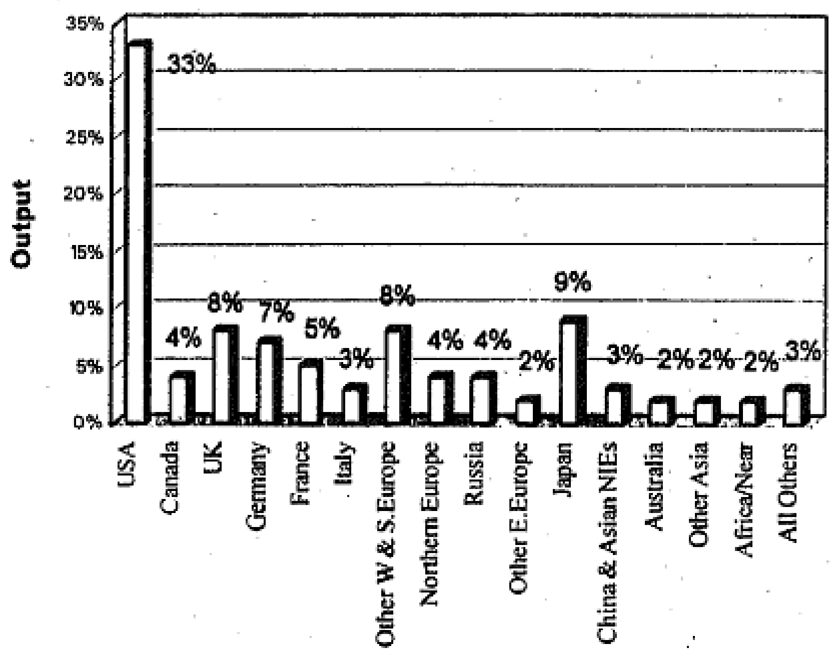



Fig.3 - An example of the use of bar diagram to depict the distribution of articles in world scientific and technical journals, by region/ country (Based on the Data at Fig. 1)

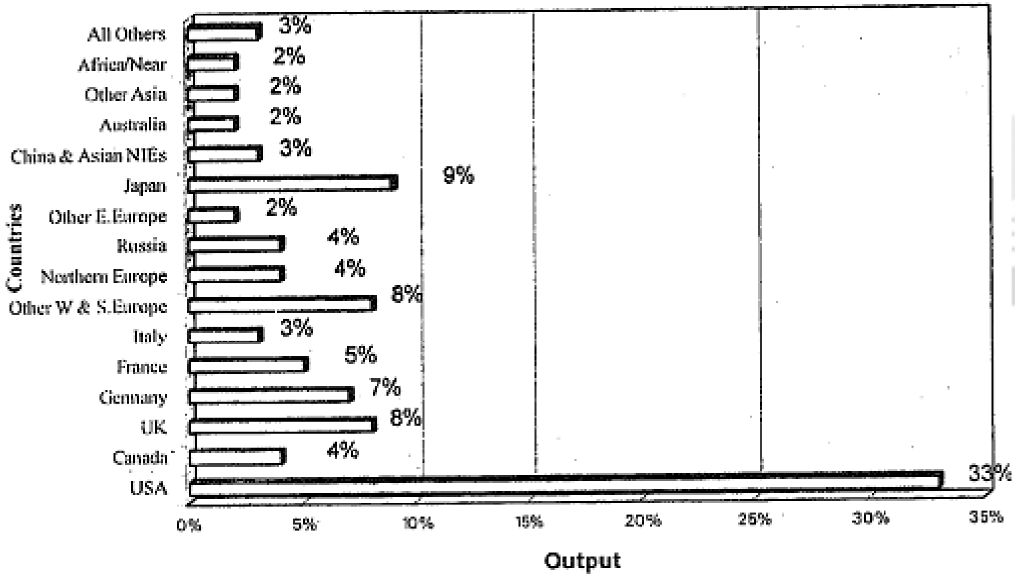
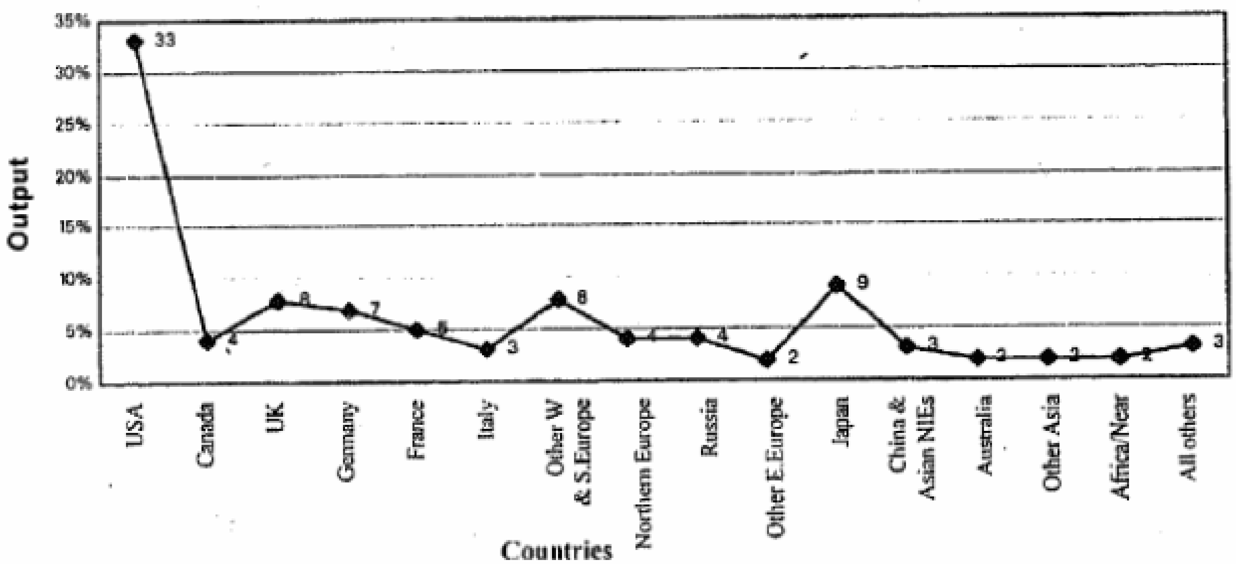



Fig.4 - An example of the use of line chart to depict the distribution of articles in world scientific and technical journals, by region/ country (Based on the Data at Fig. 1)




## 13.16 : Appendices

In constructing indicators, especially comparative literature-based indicators, you will need to classify publications into different subject categories. There are major categories, e.g., Physics, Chemistry, Mathematics and Life Sciences. These are further classified into sub-categories, e.g. condensed matter physics, high energy physics and nuclear physics. Given the nature of knowledge, it is hardly possible to have a consensus on classification of research papers and journals. Different people will classify the same set of papers (or journals) somewhat differently. Given in Appendix 1 are two widely accepted classifications, one followed by CHI Research Inc. and their clients such as the US National Science Foundation, and the other used by the European Commission. In indicators work, often classification is done at the level of the journal. Indeed, it would be a herculean task to classify each paper when one is dealing with thousands of papers. The *SCI Guide* published by the Institute for Scientific Information provides such a classification for over 4,000 journals it has listed. Some journals may be classified under more than one subject class. For example, *Journal of Molecular Spectroscopy* is listed under three subcategories, viz. spectroscopy, instrumentation and analytical sciences. *13ioresource Technology* is classified under energy, biotechnology and agriculture. Another source of classification of journals is the Ulrich database, now available in CD-ROM.

Appendix 2 lists the 50 most scientifically productive countries of the world classified by geographic region.

**Appendix 1**

Given here are the classification schemes used by two of the world's most important providers of S&T indicators, viz. the National Science Foundation (USA) and the European Commission.

Broad and fine fields for publication output data devised by CHI Research Inc.

| Broad field | Fine fields | | |
|---|---|---|---|
| Clinical medicine | Addictive disease<br>Allergy<br>Anesthesiology<br>Arthritis & rheumatism<br>Cancer<br>Cardiovascular system<br>Dentistry<br>Dermatology & veneral disease<br>Endocrinology<br>Fertility<br>Gastroenterology<br>General & internal medicine | Geriatrics<br>Hematology<br>Hygiene & public health<br>Immunology<br>Miscellaneous clinical<br>Nephrology<br>Neurology & neurosurgery<br>Obstetrics & gynecology<br>Ophthalmology<br>Orthopedics<br>Otorhinolaryngology | Pathology<br>Pediatrics<br>Pharmacology<br>Pharmacy<br>Psychiatry<br>Radiology & nuclear medicine<br>Respiratory system<br>Surgery<br>Tropical medicine<br>Urology<br>Veterinary medicine |
| Biomedical research | Anatomy & morphology<br>Biochemistry & molecular biology<br>Biomedical engineering<br>Biophysics<br>Cell biology, cytology & histology | Embryology<br>Genetics & heredity<br>General biochemical research<br>Microbiology<br>Microscopy | Miscellaneous biomedical research<br>Nutrition & dietetics<br>Parasitology<br>Physiology<br>Virology |
| Biology | Agriculture & food science<br>Botany<br>Dairy & animal science<br>Ecology | Entomology<br>General biology<br>General zoology | Marine and hydro-biology<br>Miscellaneous biology<br>Miscellaneous zoology |
| Chemistry | Analytical chemistry<br>Applied chemistry<br>General chemistry | Inorganic & nuclear chemistry<br>Organic chemistry | Physical chemistry<br>Polymers |
| Physics | Acoustics<br>Applied physics<br>Chemical physics | Fluids & plasmas<br>General physics<br>Miscellaneous physics | Nuclear & particle physics<br>Optics<br>Solid state physics |
| Earth and space sciences | Astronomy & astrophysics<br>Earth & planetary science<br>Environmental science | Geography<br>Geology<br>General engineering | Meteorology & atmospheric science<br>Oceanography & limnology<br>Metals & metallurgy |
| Engineering and technology | Aerospace technology<br>Chemical engineering<br>Civil engineering<br>Computers<br>Electrical engineering & Electronics | Industrial engineering<br>Library & information science<br>Materials science<br>Mechanical engineering | Miscellaneous engineering & technology<br>Nuclear technology<br>Operations research & management |
| Mathematics | Applied mathematics<br>General mathematics | Miscellaneous mathematics | Probability & statistics |

**Source -** CHI Research Inc., Science Indicators database; National Science Board - *Science and Engineering Indicators - 1998,* National Science Foundation, Arlington, VA, USA, 1998 (NSB 98-1), p. A-268.

| Science fields and sub-fields used in the bibliometric analysis included in the European Report on S&T Indicators. | |
|---|---|
| 1. Clinical Medicine<br>1.1 Public Health<br>1.2 Neurosciences<br>1.3 Reproduction Medicine & Geriatrics<br>1.4 General Medicine<br>1.5 Internal Medicine | 1. Physics<br>1.1 Nuclear Sciences<br>1.2 Applied Physics<br>1.3 Solid State Physics<br>1.4 Other Physics<br>1.5 Instruments |
| 2. Biomedical Research<br>2.1 Pharmacology & Pharmacy<br>2.2 Pathology<br>2.3 Research Medicine<br>2.4 Immunology | 2. Mathematics<br>2.1 Mathematics |
| 3. Biology<br>3.1 Food Science & Agriculture<br>3.2 Ecology<br>3.3 Microbiology<br>3.4 General Biology | 3. Engineering<br>3.1 Materials Science<br>3.2 Electronic Engineering<br>3.3 Mechanical Engineering, Civil & Other Engineering<br>3.4 Biotechnology<br>3.5 Informatics<br>3.6 Computer Science |
| 4. Chemistry<br>4.1 Inorganic Chemistry & Engineering<br>4.2 Analytical Chemistry<br>4.3 Physical Chemistry<br>4.4 Organic Chemistry | 4. Earth and space sciences<br>4.1 Geosciences |

**Source -** RASCI: European Commission, *Second European Report on S & T Indicators 1997*, Office for Official Publications of the European Communities, Luxembourg.

**Appendix 2**

**The 50 most active S&T countries in the world**

By Geographical criteria

**EUROPE:**

**Western Europe** : Belgium, Denmark, Germany, Greece, Spain, France, Ireland, Italy, Luxembourg, Netherlands, Austria, Portugal, Finland, Sweden, United Kingdom, Iceland, Norway, Switzerland

**Eastern and Central Europe** :Czech Republic, Slovak Republic, Hungary, Poland. Bulgaria, Romania, Russian Federation, Ukrain, Turkey, Albania

**NORTH AMERICA** : Canada, Mexico, USA

**SOUTH AMERICA** : Argentina, Brazil, Chile, Venezuela

**ASIA South East Asia** : China, Hong Kong, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, Thailand, Taiwan

**Other Asia** : India, Pakistan

**OCEANIA** : Australia, New Zealand

**Other Africa** : South Africa

**Other Mediterranean** : Cyprus, Algeria, Egypt, Morocco, Israel, Lebanon, Malta, Tunisia, Syria

**Source- European Commission :** ***Second European Report on S& T Indicators,1997- Report,*** **Office for Official Publications of the European Communities, Luxemburg, p3.**

## 13.17 : Summary

In assessing performance and usefulness of science, the development of tools like science indicators is of utmost importance. Through the use of a set of indicators, it is possible to assess and evaluate the scientific and technological activities of countries which ultimately help the planners, policy makers and decision makers. By studying several indicators it is possible to discern trends in the behaviour of certain phenomenon. The S&T indicators also help in understanding the structure of science and its impact on the S&T policies and programmes which have implication on the social and economic development of a country. Its importance has, thus, led many countries and international organizations develop S&T indicators. From the very beginning such indicators have been developed by developed countries. The OECD countries and the European Unions constitute at the Centre and the developing countries like Africa, Asia and Latin America constitute the periphery. There are various S&T indicators which have been developed and are available. Depending upon the purpose and scheme of evaluation, one or a combination may be adopted. To measure scientific productivity there are literature based indicators, patent based indicators, human resource indicators etc. Each has its positive aspects and limitations. The countries which have developed such indicators have benefited the modern science and technology. For the developing countries whose concerns are different from those of economically developed countries, it is necessary to formulate a fairly complete system of indicators which should be able to reflect the needs and requirements of the objectives of socio-economic development of such countries.

# Unit-14:Mapping of Science

## 14.0 : Objectives

In this Unit, you will learn the concept of mapping science, and get a basic understanding of the different kinds of mapping. Starting from the metaphorical use of the word mapping', you will learn the elements of both - cognitive mapping and descriptive mapping of science. After studying this unit, you would have a basic level appreciation of journal - journal intercitation mapping, document co-citation mapping, author co-citation mapping, journal co-citation mapping, co-word mapping and co-classification mapping. You will also be introduced to mapping domestic and international collaboration using author addresses in multi-authored papers and descriptive mapping of science at different levels of aggregation. The centrality of the literature of science, especially as captured by databases and in particular the citation index database of ISI, in mapping studies will be emphasized.

## 14.1 : Introduction

When we talk of mapping science, we use the word mapping metaphorically. We have all seen maps. Maps are used by geographers to provide a two-dimensional representation to scale the world we live in. In particular, maps describe the physical boundaries of land masses. They serve as tools of navigation by accurately depicting spatial relationships. In such maps two dimensions are important: the distances between locations and the directions. For example, Delhi is north-east of Mumbai and at a distance of roughly 1,400 km; USA is to the west of India and Thailand to the east. Often in an Atlas, the third dimension, viz. the altitude (or height from the mean sea level), is indicated by colors. The plains and valleys are green, the deep sea is blue and the mountains are different shades of brown.

Following this example, it is possible to construct an atlas of knowledge, by suitably positioning different fields (or specialties) in two-dimensional space. We know that mathematics is closer to physics than biology, and chemistry is related to both physics and biology. Indeed, universities offer graduate level courses in hybrid areas such as physical chemistry, chemical physics, biochemistry and molecular biology etc., and there are journals devoted to these areas. Information scientists do try to portray such semantic boundaries between the myriad areas of research. Such mapping attempts to bring out the cognitive structure of knowledge. It helps us understand the relational and structural aspects of scholarship and its dynamics.

The notion that science can be mapped was first stated by Derek de Solla Price, the author of *Science since Babylon* and *Little Science, Big Science,* in the mid-1960s. But the idea that knowledge could be represented spatially has been there even before. The very word "field' (to denote a subject or discipline) carries with it spatial connotations. Vannevar Bush, Science Advisor to President Roosevelt of the United States, used the metaphor `endless frontier' when talking about the great potential and possibilities of science. In the late 1960s, Dr. H. Bentley Glass, the well-known geneticist and past president of the American Association for the Advancement of Science, talked about researchers as "explorers of a great continent." He said, metaphorically, "We have penetrated to its margins, we have mapped the major mountain chains and rivers." Today scientists at Sandia National Laboratory in the United States are routinely displaying the large-scale structure of science as a landscape. More about it later.

There is also another kind of mapping of knowledge, which aims to provide an overview of the distribution of production of knowledge at different levels of aggregation - at the level of nations, states (within a nation), cities/towns, institutions, journals, disciplines, sub-disciplines, etc. For the sake of convenience, let us call this kind of mapping as descriptive mapping.

The construction of both kinds of maps of knowledge (or science) depends largely on counting surrogates of knowledge, viz. papers published in professional journals and cross-citations among such papers. Such bibliometric maps of science may be of help in improving information retrieval, in research management, and in science policy, and as a tool in science studies, in particular in sociological and historical studies of science.

While it is relatively easy to map science output by nation or by subject category (descriptive mapping), it is far more difficult to describe subject matter (cognitive mapping). Let us first look at different types of cognitive mapping.

## 14.2 : Cognitive Mapping

There are four principal ways of constructing cognitive maps. All of them are based on surrogates of knowledge, viz. published papers and references cited therein.

### 14.2.1 : Journal-to-journal Citation Maps

Citations are the formal, explicit linkages between papers that have particular intellectual points in common. The citation databases of the Institute for Scientific Information. USA are built around these citations - or linkages - provided by the authors. These databases list publications that have been cited and identify the sources of these citations. Derek de Solla Price, like the doyen of sociologists of science Robert Merton, was among the earliest to see the great potential of Eugene Garfield's *Science Citation Index (SCI),* when it was released in the early 1960s. He saw instantly the tremendous possibilities offered by the millions of interconnections among hundreds of thousands of scientific research papers captured annually by *SCL* He recognised the potential of an inter-journal network of citations to derive maps of science. As far back as 1965, Price wrote a paper on the networks of science. Subsequently Henry Small, Director of Research at ISI, and Eugene Garfield used these data to derive natural maps of the scientific landscape.

The use of inter journal citation-frequency rests on the assumption that it serves as an indication of the magnitude of subject relation between journals. Journals are the primary formal channels for communicating theories, methods and empirical findings to researchers, and therefore they may indicate changing research frontiers and disciplinary transitions. The network of journal journal citation links thus reflects the macro-level structure of scientific activities. Changes therein indicate the dynamics of scientific fields. Data on inter journal citation frequencies can be obtained from *Journal Citation Reports,* produced annually by ISI. Such data have been used to rank the importance of journals in different fields. Table 1 presents a cross journal citation matrix of astronomy and astrophysics journals constructed using data from *Journal Citation Reports* 1997. Such maps have been used to rank the influence of journals in selected fields.

**Table 1: Cross-journal citation matrix of astronomy and astrophysics journals**

**Cited Journals (1988-97)**

| | J1 | J2 | J3 | J4 | J5 | J6 | J7 | J8 | J9 | J10 | J11 | J12 | J13 | J14 | J15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All | 26755 | 71519 | 21159 | 11948 | 3070 | 407 | 366 | 40083 | 4796 | 14414 | 10940 | 135825 | 228945 | 269100 | 271952 |
| J1 | 8524 | 11586 | 3590 | 2029 | 510 | 46 | 5 | 70 | 1369 | 36 | 59 | 109 | 149 | 792 | 19 |
| J2 | 6209 | 26168 | 5926 | 2862 | 924 | 35 | 32 | 292 | 872 | 110 | 126 | 380 | 303 | 2190 | 27 |
| J3 | 2502 | 7704 | 5204 | 1102 | 366 | 18 | 9 | 60 | 420 | 2 | 39 | 67 | 71 | 674 | 14 |
| J4 | 1362 | 4395 | 1228 | 2445 | 228 | 13 | 0 | 0 | 421 | 3 | 4 | 2 | 51 | 222 | 5 |
| J5 | 454 | 1178 | 303 | 258 | 68 | 2 | 0 | 0 | 34 | 18 | 7 | 3 | 18 | 95 | 19 |
| J6 | 187 | 382 | 112 | 122 | 20 | 143 | 0 | 3 | 63 | 0 | 3 | 2 | 3 | 39 | 0 |
| J7 | 58 | 218 | 28 | 4 | 10 | 0 | 73 | 141 | 3 | 0 | 23 | 97 | 9 | 46 | 0 |
| J8 | 108 | 1014 | 230 | 10 | 38 | 2 | 74 | 13496 | 6 | 0 | 255 | 4023 | 62 | 159 | 33 |
| J9 | 1214 | 1403 | 587 | 493 | 64 | 9 | 0 | 0 | 690 | 0 | 6 | 2 | 13 | 73 | 2 |
| J10 | 40 | 103 | 7 | 4 | 6 | 0 | 0 | 0 | 21 | 1508 | 6 | 36 | 536 | 594 | 3 |
| J11 | 91 | 361 | 33 | 24 | 26 | 0 | 0 | 151 | 5 | 14 | 65 | 551 | 52 | 100 | 20 |
| J12 | 33 | 254 | 68 | 7 | 11 | 0 | 25 | 1268 | 3 | 20 | 495 | 11160 | 666 | 707 | 80 |
| J13 | 142 | 372 | 39 | 94 | 40 | 0 | 0 | 4 | 13 | 229 | 29 | 529 | 3374 | 2509 | 1382 |
| J14 | 1041 | 395 | 81 | 73 | 36 | 0 | 0 | 11 | 8 | 145 | 28 | 442 | 1841 | 4399 | 1119 |
| J15 | 4 | 62 | 10 | 3 | 2 | 0 | 5 | 3 | 0 | 23 | 20 | 134 | 5148 | 5966 | 8555 |

| | | (Impact Factor) | | | (Impact Factor) |
|---|---|---|---|---|---|
| J1 | Astron Astrophys | 2.362 | J8 | Phys Rev D | 3.420 |
| J2 | Astrophys J | 2.948 | J9 | Astron Astrophys Supp | 1.136 |
| J3 | MNRAS | 3.383 | J10 | Geophys Res Lett | 2.180 |
| J4 | Astron J | 2.419 | J11 | Rev Mod Phys | 12.623 |
| J5 | Annu Rev Astron Astr | 9.618 | J12 | Phys Rev Lett | 6.140 |
| J6 | Astron Lett | 0.513 | J13 | Science | 24.676 |
| J7 | Astropart Phys | 2.062 | J14 | Nature | 27.368 |
| | | | J15 | PNASc | 9.040 |

Data from *JCR* -1997

Here are some typical questions one can answer using ajournal-to-journal citation matrix:

- a) Which journals represent various disciplines or provide channels for diffusion of ideas across the disciplines?
- b) Which journals represent sub-disciplines and how are those sub-disciplines related?
- c) Which journals receive proportionately more citations?
- d) Do disciplinary and sub-disciplinary structures change over time, say converge or integrate?

Fora good example of a journal journal citation mapping study, please see "Hierarchies and clusters among communication and library and information science journals, 1977-1987" by Ronald E Rice, chapter 8 in the book *Scholarly Communication and Bibliometrics,* ed.by Christine L Borgman,1990, Newbury Park: Sage Publications, Newbury Park; pp. 138-153.

## 14.2.2 : Co-citation Maps

Co-citation coupling is used to establish a subject similarity between two documents. If papers A and B are both cited by paper C, they may be said to be related to one another, even though they do not directly quote each other. If papers A and B are both cited by many other papers, they have a stronger relationship. The more papers they are cited by, the stronger their relationship is. Co-citation maps show the extent to which two articles are cited together in other articles. The method for mapping such structures of science is known as co-citation analysis. It involves the application of clustering and scaling methods to data on the number of times earlier papers are cited together by later papers. It was in the late 1960s that Henry Small of ISI, and independently Irma Marshakova in Moscow, discovered co-citation mapping. Co-citation analysis, originally developed within ISI for exploitation of its *SCI* database, has become a well-established descriptive tool. Now the *Web ofScience* database provides coverage from the present back to 1945 of *SCI Index,* from the present back to 1956 *of Social Science Citation Index* and from the present back to 1975 of *Arts and Humanities Citation Index.* The key to taking full advantage of the wide range of information contained in the back-year files lies in the most dynamic and unique feature *of Web of Science, viz.* cited reference searching. This powerful feature enables researchers to navigate through the literature backward or forward in time by using the authors' own citations.

Garfield actively stimulated the development and use of maps of science to obtain "an instant history of science" through ISI's *Atlas of Science.* In his early work on historiography in collaboration with Sher and Torpie, Garfield Chad shown the longitudinal structure of science by exploiting citation links across time. Clustering procedures make it possible to obtain co-citation clusters which reflect the structure and content of research-front specialties and at higher (co-citation threshold) levels co-citation clusters comparable to large scientific fields or disciplines.

Here is Garfield's own description of the essence of co-citation clustering.

"When we record the references in an article, we also keep track of the pairs of papers it cited together, or co-cited. When the same pairs of papers are cited together with other papers by many authors, a cluster of research begins to form. The idea underlying this method is that co-cited articles share a common topic, subject area, or method. That is, when an author cites two papers together, he or she indicates that both papers applied to some aspect of the research he or she performed. When the same cluster of papers is co-cited by several authors, a cognitive link is established between the researches these authors perform. That is, the citing authors reveal what research area they "belong to" through their citations. By examining the titles of the citing articles, we get an idea of the cognitive content of their research concentration. That is, the citing authors themselves provide the words to define what their research area is about.

"At another level, we can also identify co-citation links between clusters. That is possible when authors cite articles included in different clusters together. Again, we assume that authors who use the same clusters of articles are engaged in similar types of research. The more frequently two clusters of articles are co-cited, the closer they are in subject area or methodology. In fact, when we generate maps of co-cited clusters, we use this relationship to determine how close or far apart clusters should be depicted in relation to each other."

The end result of co-citation clustering is a series of hierarchic maps showing the multi-dimensional landscape of science at five levels of aggregation- from the most general global level (C5) down to the smallest research front (C l). At the lowest level a research front is identified by groups of co-cited core papers - from 2 to an arbitrary maximum of 50 or 60. And for each new annual research front there is an annual count of currently published papers that cite into those core papers.

Usually, the network of relations between clusters is visualized by applying multidimensional scaling methods. Imagine taking a map of India and constructing a table showing inter-city distances. How can we achieve the reverse, viz, constructing the map from the table of distances? This can be achieved by scaling techniques, albeit approximately.

There are three kinds of co-citation analysis based on the three parts in a reference to a journal article, viz. name(s) of author(s), journal title, and further bibliographic information such as volume; year and page. Document co-citation analysis, which maps linkages among individual articles (or books) on the basis of their joint citation by later writers, uses all three parts (which are needed to uniquely define a document). Author co-citation analysis, which uses computational and graphic display techniques to produce maps of prominent authors in selected areas of scholarship, uses only the co-cited author names. Author cocitation analysis was introduced by Howard D White and Belver C Griffith, both of Drexel University in Philadelphia, in 1981. White and Griffith used *oeuvres* (or complete works) by authors, as opposed to single papers as the unit of analysis, and the co-citation of oeuvres as the variable that indicates authors' distance from each other in maps. In journal co-citation analysis, the journal is the unit of analysis and one uses only the journal title. in references.
Figure 1 is an author co-citation map showing the intellectual base o scientometrics. Garfield and Derek De Solla Price, as expected, are the centra figures. This map was constructed by the Swedish information scientist ©l1 Persson using first author co-citations among references appended to the more than 1060 papers published in *Scientometrics* since its inception in 1978 till 1999.

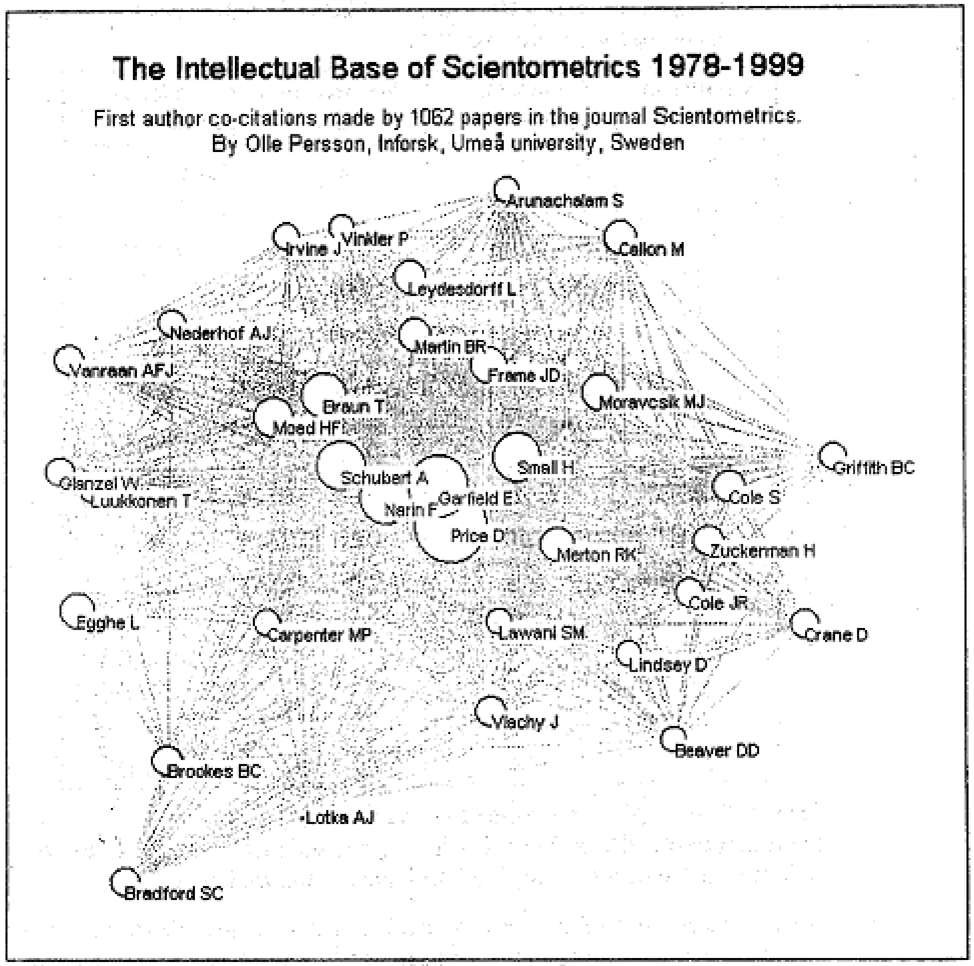



**Fig. 1: Author to-citation map showing the intellectual base of scientometrics**

The most cited authors, or to be more exact the most cited first authors, are listed in Table 2. At the top we find Price and Garfield who are the founders of the field of scientometrics. Price is cited in every third paper and Garfield in every fourth. Below them there are authors having made substantial contributions to the field and most of them have already received the "Derek Price Award", which is indicated with an asterisk. [This award is given by the journal *Scientometrics* to the most outstanding contributors to the field *of* scientometrics. The awardees are selected by an electoral college consisting of members of the editorial board of *Scientometrics]*. Such lists have great validity. To discover the structure of this research field we can count the number of times the listed authors appear together in the same reference lists of papers. This technique is called co-citation analysis and assumes that authors that are frequently co-cited are more related in terms of specialization than authors that seldom or never occur in the same reference list.

**2:Table 2: Most cited first author co-citation among references appended to the more than 1060 papers published in Scientometrics since its inception in 1978 till 1999.**

| Number of citations | Cited first-author | Number of citations | Cited first-author |
|---|---|---|---|
| 343 | Price D | 56 | Zuckerman H |
| 245 | Garfield E | 55 | Vlachy J |
| 178 | Narin F | 55 | Brookes BC |
| 155 | Schubert A | 50 | Glanzel W |
| 152 | Small H | 49 | Crane D |
| 124 | Braun T | 49 | Bradford SC |
| 96 | Moed HF | 48 | Irvine J |
| 83 | Frame JD | 48 | Carpenter MP |
| 82 | Moravcsik MJ | 47 | Griffith BC |
| 78 | Egghe L | 46 | Lotka AJ |

| | | | |
|---|---|---|---|
| 73 | Martin BR | 45 | Nederhof AJ |
| 71 | Callon M | 43 | Vinkler P |
| 70 | Leydesdorff L | 43 | Lawani SM |
| 69 | Merton RK | 42 | Luukkonen T |
| 61 | Cole S | 41 | Arunachalam S |
| 56 | Vanraan AFJ | 40 | Lindsey D |
| 56 | Cole JR | 40 | Beaver DD |

For a good example of a document co-citation study, please refer to "A co-citation study of AIDS research" by Henry Small and Edwin Greenlee, chapter 10 in the book edited by Christine L Borgman mentioned above.

For a good example of an author co-citation analysis study, please see "An author co-citation analysis of two research traditions: technology transfer and the diffusion of innovations" by Everette M Rogers and Charlotte A Cottrill, chapter 9 in the book Scholarly Communication and Bibliometrics, edited by Christine L Borgman (Sage Publications, Newbury Park), 1990, pp. 157-165.

The basic assumption underlying the use of all types of bibliometric maps is that the frequency of element co-occurrence directly reflects subject relatedness. The element may be references cited, keywords used or even classification terms. On the other hand, co-occurrence of addresses - in multi-authored papers can be used to map domestic and international collaboration in science.

*The Sandia Map* - Scientists at Sandia National Laboratories in New Mexico, USA, have extended Small's co-citation mapping by adding visualization. In their programme on "computing for research", they try to understand the inner nature of things. They have a computer that reads millions of scholarly papers and creates a digital landscape that helps visualise the state of scientific research. They have developed a powerful knowledge management tool called *Vinsight* that can be used for knowledge mining and discovery through interaction. In collaboration with Henry Small and David Pendlebury of ISI, they have used this software to plot 3D mountains and valleys that represent peaks and troughs of interest and connecting ridges between disciplines.

Every year, Sandia scientists measure technical advances by the more than five trillion printed words in scientific, technical and trade journals. If you read at the speed of 200 words per minute, you will need 250 years to read a year's worth of the top journals. How do we handle such voluminous information and how do we take advantage of all the knowledge that is out there? Sandia scientists use powerful supercomputers to build knowledge maps out of these large volumes of papers that let one navigate through virtually all scientific research over two decades.

In one of the early maps, they looked at over 43 million papers world wide, written during the 16 years 1980-1995. Using the most highly cited 37,000 papers, Sandia scientists constructed the landscape showing similarity (relative distance) and concentration (elevation) of papers. In essence, what they have done is to use the source information contained in the data to build maps that explicitly show the evolution and structure of knowledge. In other words, they have organized data in a visual way in order to gain insights much more quickly.

Science is produced through a network of interactions among researchers. This network and science evolve together. It is this co-evolution that drives the idea that we can portray science as an evolving landscape and can learn much from what we observe. The landscape shows the dynamics of the evolution of science, and ultimately may help us improve our ability to invest in leading areas.

There is also the reverse or mirror image of co-citation coupling. Called `bibliographic coupling', it links two papers that cite the same articles, so that if papers A and B both cite paper C, they may be said to be related, even though they do not directly cite each other. The more papers they both cite, the stronger their relationship is.

### 14.2.3 : Co-word Maps

Analogous to co-citation analysis, counts of word co-occurrences (co-word frequencies) are used to construct a co-word structure representing research themes in a discipline and their inter-relations, thus capturing relevant features of the cognitive structure of scientific (sub) specialties. The idea is if two or more articles use several keywords in common, then those articles have something in common. Co-word analysis was developed largely in France, researchers at the Centre for the Sociology of Innovation at the Superior School of Mines and the Institute for Scientific and Technical Information, both located in Paris, in the early 1980s. Unlike co-citation analysis, it can provide a quick idea of the actual content of the research topic dealt with in the literature. It is especially useful in subfields and research fronts which do not conform to the normal citation patterns of scientific literature. One may not cite many references, but one has to necessarily use words in writing a research paper. French co-word analysts have developed a computerized system for co-word mapping, called LEXIMAPPE, which is applied in sociological studies of science as well as for science policy purposes. In this method, one starts with the counting of association of keywords and arranging them in decreasing order of their strength. For example, if the keywords A and B occur together 100 times in a set of 200 articles, and keywords B and C occur together 80 times, then the A-B co-occurrence has greater strength than the B-C co-occurrence. Then one selects words having most links with several other words to form clusters: These clusters form the basis for further steps which finally result in a two-dimensional map of a chosen part of the literature. Often co-word analysis is used to follow the evolution of a research front over time and to discern changes in emphasis in research.

### 14.2.4 : Co-classification Maps

Co-classification maps are based on the analysis of the co-occurrence of subject-classification terms. The classification terms (words or phrases) mostly represent cognitive elements which relate to scientific topics, specialties, or fields. Like co-word and co-citation methods, this approach hinges on the assumption that these items reflect the structure and degree of relatedness in research publications. Co-classification analysis is simpler than co-citation and co-word analyses. It can quickly provide a useful overview of a field. It can offer visualisation of the gross or coarse structure of a research field. But, it may not work welt with multidisciplinary fields, as most classifications are still discipline-oriented.

For an example of co-classification mapping, please see Rodosvet Todorov's article "Co-classification analysis for science *mapping:* An example from superconductivity" in *Science and Technology Indicators*, edited by A F .J Van Raan, A J Nederhof and H F Moed 1989,DSWO Press, Leiden University, pp. 263-270.

When does one use co-citation, co-word and co-classification analysis? Indeed, there is an ongoing rivalry between the supporters of co-citation analysis and co-word analysis. But as a rough guide, one can use co-classification analysis as a first step to gain insight into the coarse structure of knowledge. One can use, at the meso.level, journal co-citation and author co-citation analysis to analyse specific subfields or specialties. For fine structure mapping or micro-level analysis of a research front, one can use either document co-citation analysis or co-word analysis, depending on the characteristics of the literature of the research front.

## 14.3 : Descriptive Mapping

Much of the work in cognitive mapping is done in the United States and Europe. Most mapping work in developing countries is of the descriptive kind. In descriptive mapping one is concerned with quantifying different aspects of scientific activity at different levels of aggregation. Here are some typical questions:

1) What is the volume of research carried out in India in the life sciences?

2) Which are the prominent institutions carrying out life science research in India?

3) Where (in which journals) do Indian life scientists publish their work?

4) What percentage of papers of Indian life scientists appear in journals of impact factor higher than 4.0?

5) How often do Indian life scientists publish their work in British (or American) journals?

6) Which are the Indian cities and states where research in life sciences flourish?

These questions can be answered by analyzing data from *Biological Abstracts.* Please see the paper "Mapping life sciences research in India: A profile based *on BIOSIS* 1992-1994" by Subbiah Arunachalam.

Let us look at one such exercise so that we can understand what is involved in such mapping. Imagine you want to map fish and aquaculture research in India. How will you go about it? First, you have to identify all papers published from India in this area in a chosen period. Obviously, one cannot correspond with all the authors. There are too many of them. Besides, how will one identify all the Indian authors?

One can look up annual reports of all Indian organizations likely to have published papers in the fields of fish and aquaculture research. One may miss papers written by individual authors whose institutions do not publish an annual report or who are not associated with any institution. Some institutions may not publish annual reports or may not list all papers published by their scientists. One can scan all journals likely to publish papers in this field. That is virtually impossible, as no library in India can afford to have a comprehensive collection of journals.

One can choose the right databases that would index research papers in fish science and aquaculture and look for papers from Indian addresses. Fish and aquaculture literature is indexed, among other databases, in *Aquatic Science and Fisheries Abstracts* published by Cambridge Scientific Abstracts, *CAB Abstracts* published by CABI, *Biological Abstracts* published by BIOSIS, *Science Citation Index, Biochemistry and Biophysics Citation Index* and *Biotechnology Citation Index, all* published by ISI. These databases are available in several forms: print on paper, CD-ROM, through online vendors such as STN, and on the World Wide Web. Both the coverage and the kind of information given differ from one database to the other and even between different formats of the same database. If one is working with the CD-ROM or the Web version of a database, one can download all papers from India published in a given year by giving appropriate commands [`India' in address field and '1998' in publication year field, etc.]. Also, one need not download all the information given for each entry. One can choose what one wants. For example, one can download names and initials of authors, institution, city, journal title (with volume, year, page), document type, etc. and omit the abstract. Given in Table 3 are examples of some typical records from five different databases. The formats differ from database to database.

**Table 3: Sample entries of Indian publications downloaded from five databases**

**Downloaded From *ASFA* (Web edition)**
[in full, before editing]

TI: Title
Poisonous effects of Euphorbia tirucalli on fish

AU: Author

Kamat, DV; Muthe, PT
AF: Author Affiliation

Shri C.B. Patel Res. Cent., Chem. and Biol. Sci., 2nd Flr., Mithibai Coll. Bldg., Vile Parle (West), Bombay - 400 056, India

SO: Source
Journal of Animal Morphology and Physiology [J. ANIM. MORPHOL. PHYSIOL.], vol. 42, no. 1-2, pp. 65-68, Dec 1995

IS: ISSN
0021-8804

AB: Abstract

Human beings have been using plants for catching fish from time immemorial. Some part of the plant which is poisonous to fish like seed, flower, leaf or bark is thrown into water to stupify the fish (Wallis, 1985). Maharashtra State Gazetteers (1961) has reported fourteen such plants which are poisonous to fish. The plant Euphorbia tirucalli, spurge, is from the family Euphorbiaceae. It is poisonous to fish and is used to stupify fish on the west coast of Maharashtra. The present paper describes the poisonous effect of the plant on fish. An attempt is made to, isolate the active component of plant and to study the effect of the poisonous component on fish and to evaluate LD sub (50). An active component was identified as 4-deoxy phorbol, a terpene.

LA: Language
English

SL: Summary Language
English

PY: Publication Year 1995

PT: Publication Type
Journal Article

DE: Descriptors

Euphorbia tirucalli; poisoning; 4-deoxy phorbol; plant extracts; Pisces; thin-layer chromatography; toxicity tests; terpenes; metabolites; biological poisons; catching methods; fish poisoning; Euphorbia tirucalli; Pisces

ER: Environmental Regime
Freshwater

TR: ASFA Input Center Number
CS9704386

CL: Classification
X 24172 Plants; Q 101563 Fishing gear and methods; Q5 01504

Effects on organisms

SF: Subfile

Toxicology Abstracts; ASFA I: Biological Sciences & Living Resources; ASFA 3: Aquatic Pollution & Environmental Quality

AN: Accession Number

4005566

**Downloaded From BIOSIS - *Biological Abstracts***

[Only fields necessary for analysis are selected]

TI: Satiation time and satiation amount of food of a hill stream fish Barilius bakeri Day. AU: JACOB-L; SURYANARAYANAN-H

CS: Dep. Aquatic Biol. Fisheries, Univ. Kerala, Trivandrum-695 007, Kerala, India SO: JOURNAL OF ANIMAL MORPHOLOGY AND PHYSIOLOGY 37(1-2): 23-28 PY: 1990 LA: English DE: DIET; STOMACH CAPACITY; PREDATORY ABILITY

ST: Animals; Chordates; Vertebrates; Nonhuman vertebrates; Fish

**Downloaded From SCI**

[Only fields necessary for analysis are selected]

Authors: Ramana-YV **Pandey-AK** Singh-S

Abbrev. source: BULL ENVIRON CONTAM TOXICOL 1992, Vol 48, Iss 6, pp 907- 913

Title: Dimethoate Toxicity to Gestational Embryonic Ovary of a Live Bearing Fish, Lebistes-Reticulatus

Addresses: JAYDEE-AGROCHEM-PRIVATE-LTD, RD 8, VK IND AREA, JAIPUR 302013, INDIA-VIKRAM-UNIV, SCH STUDIES ZOOL, ENDOCRINOL & ENVIORNM PHYSIOL UNIT, UJJAIN 456010, MADHYA-PRADESH, INDIA

Document type: Article

**Downloaded From *BBCI***

[all fields, before editing]

TI: Inhibitory effect of sodium chloride on the acid production of lactobacilli. AU: Abraham-TJ

AD: Microbiology Unit, Fisheries College and Research Institute, Tamil Nadu Veterinary and Animal Sciences University, Tuticorin 628 008, Tamil Nadu, India.

SO: Cheiron. 1993, 22: 4, 167-168; 7 ref. PY: 1993

LA: English

AB: This paper reports the inhibitory effect of salt on the acid production of lactobacilli dairy starter cultures in liquid medium. The results revealed that a salt concentration of up to 8% could be used for higher acid production by Lactobacillus.

DE: starters-; cultured-products; acids-; production-; inhibition-; chlorides-; sodium-chloride

OD: Lactobacillus-acidophilus; Lactobacillus-plantarum; Lactobacillus-RN: 7647-14-5

BT: Lactobacillus; Lactobacillaceae; Firmicutes; bacteria; prokaryotes; Gram-positive-bacteria

CC: ZZ390; 00010; QQI20

CD: General-Microbiology; Milk-and-Dairy-Produce; Microbiology-of-Food-Processing

PT: Journal-article IS: 0379-542X

UD: 950516

AN: 950400258

**Downloaded From *CAB Abstracts***
[only fields necessary. for analysis are selected]

TI: On Rhabdochona guptii n.sp. from the intestines of Cyprinus carpio spicularis, a fresh water fish from Hokarsar (wetland) in Kashmir.

AU: Chishti-MZ; Bakshi-S

AD: Post-graduate Department of Zoology, University of Kashmir, Srinagar- 190006, Jammu and Kashmir, India.

SO: Indian-Journal-of-Parasitology. 1992, 16: 2, 161-163; 1 ref. PY: 1992

LA: English

CC: LL820; ZZ380

PT: Journal-article

Fields Selected for Analysis:

TI: Title

AU: Author
AD: Address
SO: Source
PY: Publication Year LA: Language
PT: Publication Type

Next comes the task of merging the data from the different databases. This is the time to weed out duplicates [the same paper may be indexed in more than one database] and to standardise the format of all entries. Remember databases do contain erroneous information. For example, addresses of institutions in towns in Pakistan or Nepal may include India as the country name. Other problems one may face include wrong spelling of names of institutions [e.g., Punjab University in Chandigarh may be listed as Punjab University, which is in Lahore, Pakistan], change of names of institutions [e.g., Sagar University in Saugar has been renamed as Dr. Hari Singh Gaur Vishwavidyalaya, and Meerut University has been renamed Chaudhary Charan Singh University], and change of names of towns [e.g., Cochin in Kerala is now known as Kochi, and Madras has been renamed as Chennai and Bombay is now Mumbai]. Then one allocates cities/ towns to states and institutions to larger bodies [e.g., Central Food Technological Research Institute, Mysore, is part of CSIR, and Central Institute of Fisheries Technology, Kochi, belongs to ICAR]. Journal titles may not be abbreviated the same way in different databases. Some databases such as *SCI* and *Mathsci* give addresses of all authors, whereas *Zentralblatt* far ***Mathematik*** (Germany) does not give any address. Most databases give the address of only the first authors of papers. Thus, if one wants to map collaboration in research, only citation index databases of ISI and *Mathsci* can be used.

Coming back to fish research, in India, here are some findings from a study carried out at the M. S. Swaminathan Research Foundation, Chennai. Between 1994 and 1999, (both years included), India accounted for 2,450 papers, as seen from all the five databases mentioned earlier. This comes to roughly 5.5% of the papers published in the entire world. Of these,, about 2,040 have been published in about 300 journals. More than 1,400 papers were published in 113 Indian journals, most of them having an impact factor of 0.0, meaning they aie not indexed in *SC'1*. Indian fish researchers publish a large number of papers in British and American journals. Kochi, Kerala, leads the list of cities in number of papers, followed by Chennai, Mumbai and Mangalore. More than 400 institutions in India have contributed to the literature in this field. Some universities such as Banaras Hindu University, Madurai Kamaraj University, Osmania University and Kalyani University have a much better publication record than government institutions devoted to fisheries research, not necessarily in terms of numbers of papers but in terms of impact factors of journals in which their scientists publish. Sri Sathya Sai College in Bhopal has only two papers, but both in a high impact journal, viz. *FASEB Journal:* Most Indian Council of Agricultural Research institutes devoted to fisheries research publish most of their work in non-SCI journals of low visibility. However, they may satisfy the needs of the local clientele. Among the 50 journals which published the largest number of papers from India, only two had an impact factor greater than unity. About 60 journals were identified as almost entirely devoted to fisheries/ aquaculture studies, but all of them are low-impact journals. Only four of these journals have an impact factor of between 1.0 and 1.7, as seen from *Journal Citation Reports 1998.* That overall Indian work in fish research has appeared in low-impact journals shows that very little work is carried out in new biology areas such as biochemistry, molecular biology and biophysics.

Considerable work has been done in mapping collaboration among researchers. In a typical study, one looks at the addresses of all the authors of individual papers in a set of papers. Obviously, in single author papers there is no possibility of collaboration. In multi author papers, there are several possibilities:

a) All authors from the same department. Collaboration between authors.
b) Authors from different departments of the same institution. Inter-departmental collaboration.
c) Authors from different institutions in the same city. Inter-institutional collaboration.
d) Authors from different institutions in the same country. Inter-institutional and inter - city collaboration.
e) Authors from different institutions in the same state. Inter-institutional and domestic

collaboration.

f) Authors from different countries. International collaboration.

Many studies deal with international collaboration in science. Especially after the advent of the Internet and the email, it has become much easier for people to work with collaborators from virtually anywhere in the world. Besides, organizations such as the European Union encourage and promote international collaboration among scientists. In fact, the percentage of papers resulting from international collaboration is rising fast virtually in all fields. However, in developing countries like India, the extent of international collaboration is much lower than that prevailing in the advanced countries.

For an example of a study on mapping collaboration, please see the paper "Science in a small country at a time of globalization: domestic and international collaboration in new biology research in Israel" by Arunachalam, S. and boss, J., *Journal of Information Science*, *26(1) 2000,* pp. *39-49.*

Funding for such mapping studies in India became available only in the early 1990s, when the National Science and Technology Management Information System unit of the Department of Science and Technology funded a project on mapping India's contribution to different fields as reflected by major databases. This project was carried out by Subbiah Arunachalam at the Central Electrochemical Research Institute, Karaikudi, and later at the M. S. Swaminathan Research Foundation in Chennai. Arunachalam and colleagues mapped India's contribution to the literature of science as a whole (based on *SCI),* mathematics (based on *Mathsci* database of the American Mathematical Society), physics (based on INSPEC - *Physics Abstracts),* medicine (based on *Medline),* and materials science (based on *Materials Science Citation Index).* Subsequently the National Information System for Science and Technology, Department of Scientific and Industrial Research, funded a number of similar projects, to quantify India's contribution to agriculture (based on *CAB Abstracts),* biology (based on *Biological Abstracts),* chemistry (based on *Chemical Abstracts),* engineering (based on *Engineering Index),* etc. Findings of some of these studies have been published in different issues of Information Today and Tomorrow.

One of the earliest to map science in the developing world was Eugene Garfield. In his Magnus Pyke Science Policy Foundation Lecture delivered at London in 1982, he referred to India as the Third World's leading performer of science and showed how India differed from other mainstream developing countries. Unlike in other Third World countries, clinical and biomedical research is not the most active area of research in India. Also, Indian science is clustered in research fields that other Third World scientists have not significantly penetrated - biochemistry, physics, chemistry and organometallic chemistry.

In a talk on mapping the world of science delivered at the 150'" Anniversary Meeting of the American Association for the Advancement of Science, held at Philadelphia, on February 14, 1998, Eugene Garfield showed the great decline in the percent share of non-English language papers in the scientific journal literature. Figure 2, based on data from *SCI*, shows how. the non-English European language representation in the *SCI* has changed over the past 20 years for seven languages. For example, German language has dropped. from about 6% to 1.5%. French has dropped from 4.5% to 1%. He had also shown the rapid pace at which the number and percent share of Chinese publications had increased.

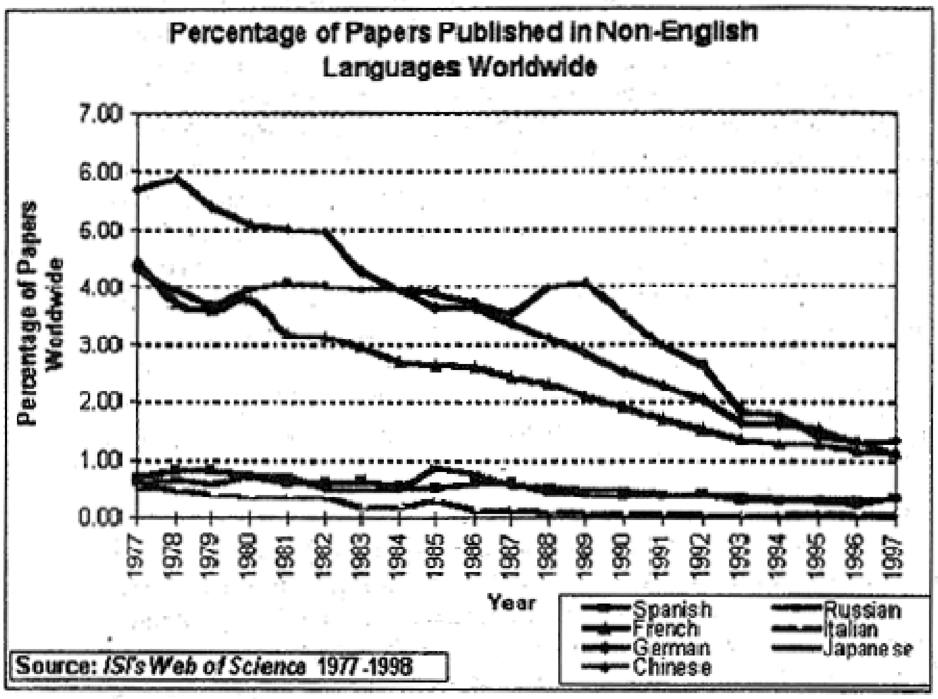



Fig. 2: Shows decline in the percent share of non-English language papers in the scientific journal literature

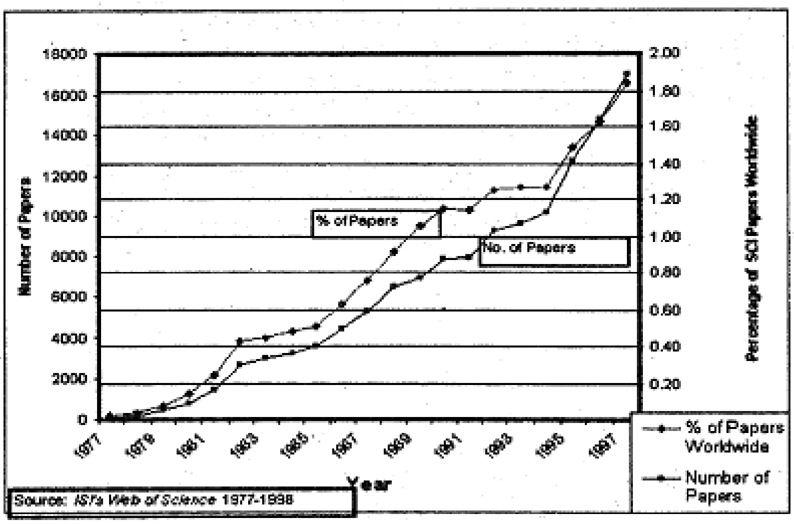



Fig. 3: Shows that from a tiny beginning in 1977 to 17,000 articles in 1997, the PRC now accounts for 2% of world output, most of which is published in the West. There are probably another 10-15,000 articles per year published in journals not included in the SCI.

We will close this Unit with one good example of mapping scientific research in India based on two years of *SCI*, 1990 and 1994, carried out by Aparna Basu and P S Nagpaul. This is one of the most detailed analyses of scientific research carried out in India, albeit using very limited data. But students interested in pursuing mapping can learn many techniques from this report. In particular, they have classified states on the basis of both output and impact factor: Maharashtra, Delhi, Karnatak and Andhra Pradesh; have scored better than national average both in output and in impact, whereas states like Gujarat and Kerala have fared poorer than national average on both output and impact. They used state-of-the-art techniques of data analysis and visualization in this study.

## 14.4 : Summary

Mapping Science is an area which is one of the applications of scientometric methods and tools. Studies and research on this area are being conducted in many countries to identify different facts of science such as structure of various disciplines and sub disciplines and their interrelationships, the scientists and their contributions in these disciplines, the important journals and authors contributing to these fields etc. The different techniques such as journal journal matrix, co-citation, co words and co classification techniques are being used to map science. A few software packages have also been developed to facilitate such studies. The analysis of contribution reflects the state-of-the-art position of different countries. Development of further techniques and tools are attempted to arrive at more and more accurate methodology for mapping science through scientometrics methods.