

Lecture to be delivered in Mexico City at the 4th Laboratory Indicative on Science & Technology at CONACYT, Mexico DF July 12-16, 1999-07-16

**For What Purpose are the Bibliometric Indicators and
How Should They Work**

By

Subir K Sen

Department of Library Science,
University of Calcutta, Asutosh Building,
Collage Street, Calcutta – 700073, India

Fax: +91 33 5512180

e-mail: subir_s@hotmail.com

After a short listing of different types of bibliometric indicators, the paper examines the viability and applicability of them. Some new bibliometric indicators have been suggested. The need for some standardization in data collection and manipulating has been stressed.

This is a preliminary draft. In the delivered talk on 12 Jul 1999 a thoroughly changed version was given. The modified version is in rough draft. Can be submitted afterwards.

Address for communication: Subir K Sen
85, (old)Devi Nivas Road
Motijhil, Calcutta-700074

Bibliometric indicators

Introduction

Among many definitions of bibliometrics one definition is that bibliometrics is a quantitative assessment of man's cultural progress including science and technology as may be revealed through bibliographic data [1]. Bibliographic data are those which can be collected, derived or deciphered from different parameters as can be assigned to a document. A document can provide two types of information –one is topical information or the so called thought content for which a reader studies a document. The other is the set of peripheral information which may be used for document description or which may be derived from the document or which may be assigned to the document from some authoritative source to describe or designate the document. These may include the name(s) of the author(s), number of pages, number of words in a part of the document, the subject classification number, the bibliographic references, citation etc.

Bibliometric indicators are based on bibliographic parameters or bibliographic features as some authors have said. They are a set of bibliometric parameters. They are the results of the need for objective and easily manipulable measures of scientific and technological activities and output. But there are now bibliometric indicators for the social sciences and the humanities also.

Most of the bibliometric indicators are arbitrary and artificially provided. They have little connection with any theoretical background or understanding of the underlying process. Not all the bibliometric indicators can be applied universally. They are contextual and some time highly specific.

Bibliometric Indicators

Very broadly defined, a bibliometric indicator is a device based on some information mechanism (usually bibliographic information) and is a conceptual tool for facilitating futuristic projection and assessment of existing state and status of an intellectual activity. In narrow specific sense a bibliometric indicator is a measure or an index or a statistic (preferably objective) to the impact or quantity of publications as documentary products [2]. These are related to literature indicators, publication indicators, science indicators, etc. Ley desdorff considers a bibliometric indicator as anything that might count about text [3]. Diodato used the definition that they are measures providing information about the nature of a subject [4].

Direct bibliometric indicators:

Direct indicators are those which use the bibliographic data available in a straightforward way from the documents. These are:

1. The number of authors per paper or the collaborators.
2. The no. of pages or no. of lines in a paper or a document
3. The proportion of the text matter and the supporting matters and the illustrative matters. In the text matter we can consider the written text from introduction to the conclusions. In the peripheral or supporting matter we consider the abstract orac knowledgement, the appendices and the list of references. In the illustrative matter we consider the tables, graphs, charts etc.
4. The no. of references or the reference size
5. References age distribution.

All such quantitative data as are directly available from the document.

Derived Indicators:

Derived indicators are those which can not be calculated directly from the documents but are to be prepared or calculated after some manipulation using the features and items implicit in the documents. These include:

1. Citation counts and all the indicators derived from citation data together with co-citational indicators.
2. Indicators calculated from the word frequency counts in the documents and their derivatives together with indicator based on co-word analysis.
3. Subject categorization of the micro-documents.
4. All the indicators based on ranking procedure of journals, countries, authors, etc based on productivity counts, reference count, citation counts, etc.

Assigned indicators:

Such indicators are somewhat extraneous and are attributed by other based on bibliographic features or assessment of thought contents of so called qualities of the documents or bibliographic items. Some of these are:

1. Indicators based on peer judgement.
2. Some of the indicators based on use of documents (these may be calculated from library lending data, document copying and supplying data, number of references, etc)
3. Indicators based on analysis of scattering
4. Subject classification of the documents

Non-bibliometric indicators:

Such indicators are based on data which are not available or can not be derived from the document description or the documents. They are not at all bibliographic items as such. They are not also assigned characteristics based on some features or aspects of the documents. Library use of documents, records of documents delivery from a documentation center, number of journals published in a country, technology transfer, per-capita-expenditure on research, gross expenditure on research are some of the items which are non-bibliometric but can be used to produce science and technology indicators.

Mixed Indicators

There can be indicators which are produced by composing both bibliographic and non-bibliographic items. Vinkler [2] made a survey of bibliometric indicators and attempted to classify them. He primarily categorized the indicators as publication indicators and citation indicators. He also considered the possible types as simple (which are single characteristic data without any standard), specific (which are characteristic data projected to other characteristics data), balance (which are one type of characteristic data related to another type of characteristic data), distribution (which are based on distribution or share or proportion of a set of characteristic data in a class of the same type of data sets), relative (these are characteristic data against a background of some absolute standard). Each of these five can represent either quantity or impact or quantity and impact. Thus, Vinkler's typology has 15 types of indicators.

He has also given a set of different levels of reference standards of bibliometric indicators calling them micro, meso and macro. Each of these can be related to different types of reference standards such as organization, thematic and publication. Vinkler's table of type and level of reference standards of bibliometric indicators is worth reproducing here:

Level of reference standards

Type of reference

Standards:	Micro	Meso	Macro
Organization	Person, institute,	Group of institutes, team, department	Countries, gripus of contries, world
Thematic	Project research	Subject field, subfield	Research discipline academic discipline
Publication	One paper	Group of papers	All papers on a subject, all papers from an institution etc

Vinkler lists 46 bibliometric indicators 7 of the mass publication indicators with one each of simple quantity, simple impact, distribution quantity indicators and 4 as simple impact and quantity indicators. Among the citation indicators his list has two simple impact, four simple impact and quantity, nine paper specific impact, three author specific impact, two specific impact and quantity, five balance impact, eight distribution impact and six relative impact indicators. Somebody should update Vinkler's list.

Purpose of bibliometric indicators:

Many of the bibliometric indicators are just intellectual exercises. They can't be used purposefully. Some of the indicators based on Shannon's information entropy are of this nature.

Many of bibliometric indicators are Contextual. The ratio of articles published in national journals and the foreign journals may give an idea of internationally or islandness of research activities of a developing country

Accumulation Versus Cumulation

Science and Technology are said to be cumulative-meaning that the new knowledge is based on immediately past knowledge thus replacing older information by new information. Because of this, the S&T information obsolesces sooner or later. On the other hand non-science literature or information is piled up and goes on accumulating. No amount of drama or poetry written after Shakespeare or Kalidasa can replace their writings or make them obsolete. However there can't be any field of subject where literature or information is absolutely cumulative and there is no subject except for creative literature which is absolutely accumulative. We have a whole spectrum from 100% accumulative to 100% cumulative. Cumulation is most evident at the research fronts of high metabolic subjects. We shall propose here some indicators for measuring cumulateness of a group of articles.

From static to dynamic indicators:

Citation indicators and many other bibliometric indicators are usually static. We require dynamic indicators if we have to exploit the changing nature of the bibliometric data. The indicators based on citation indexes are fixed. They may give a good picture for American science and technology but not for other countries. If we want to get an insight for a developing country the data-base must be thoroughly modified. Many of the indicators are data-base dependent. The same indicators may not apply to different data-bases. Only dynamic indicators can overcome this problem.

Steen in a recent article has written that there is a certain tension in the use of indicators. The process is not mechanistic, but a combination of both rational and irrational arguments. Indicators, despite their limitations, are taken into account in the political processes of allocation and resourcing by governments who take S&T seriously. Thus, he concluded, indicators are becoming more political [5].

One of the areas of application of bibliometrics and bibliometric indicators is ranking of periodicals. Ranking of periodicals is done for finding out the core periodicals in a particular subject especially for identifying journals relevant to a frontier research area which has recently emerged. This is usually done by using Bradford's Law of Scattering. Another purpose is to find out the core journals according to impact or visibility. This is done by citation counts either using a data-base like SCI's journal citation report or by creating a data-base of citation from review articles or from review articles or from a set of journals.

There is again a third method of counting the references appended to a set of articles in a set of documents and then rank the journals on reference counts. These documents may be journals, theses or books.

Nobody has so far shown either theoretically or on the basis of extensive empirical work whether these three approaches can lead to the same set of core journals in a subject. It is not always understood why such lists are produced. Are they used for resource building in libraries or are they used for making the end-users aware of extent to which these lists are practically used. In case of journal selection for a library the selection should depend on the potential use by the clients and on the basis of possibility of resource sharing in the geographical locality of the library and availability of funds. In this a user survey or use survey may be most crucial. Even for quality judgement of the journals an expert opinion-poll may be very important. The latter two methods do not depend upon any bibliometric parameter. We should probably use both bibliometric and non-bibliometric methods and parameters for this purpose. It is most unlikely that all the methods noted above and their different variants would give usable compatible results. What indicators should we use when there would be mis-match among the ranked lists derived through various methods, to arrive at a suitable list.

Sengupta [6] suggested three parameters for producing ranked lists. They are:

$$\frac{D}{A} = \frac{\text{The total no. of citation in favor of a journal from the source journals during a particular year}}{\text{The total no. of articles published in a journal during a particular year.}}$$

$$\frac{C}{A} = \frac{\text{The total no. of words published in a journal during a particular year}}{\text{The total no. of articles published in a journal during a particular year.}}$$

$$\frac{D}{C} = \frac{\text{The total no. of citation in favor of a journal from the source journals during a particular year}}{\text{The total no. of words published in a journal during a particular year}}$$

Rank lists based on them show that two of the lists produce more or less equivalent results but the third falls far apart. There is no rational choice except to follow a democratic norm with chance of being totally misled.

Conclusion

One reason for the interest in bibliometric indicators is to lessen or overcome the subjective diversion and unreliability of assessment by individual peers and experts. Such assessments are required at the individual level, at the institutional level, at the national level and at the level of supporting or non-supporting a research front. These assessments are needed in case of an individual for employment, promotion for project funding, for awards and rewards, for knowing about the status of a scholar whose expert counseling may be important towards a social or economic goal, etc.

At the institutional level one needs to know the R&D capabilities and performance standards already in existence for an institution. It is also important to know the different foci of research areas, degree of inter-disciplinarity or subject specificity.

At the national level the need is to have an information base and indication for the guidance to policy matter. Whether to give support and spend money for a particular research area, say, super-conductivity, or cold-fusion or inter-galactic space rockets. The international bodies, the national govts, the S&T leaders and even the journalists and tax payers are interested in knowing the status of performance in science and technology and in other spheres.

Bibliometric indicators are expected to provide keys to all these. But in many cases bibliometric parameters are insufficient. One needs to consider non-bibliometric items, such as, expenditure or income in terms of money, the time spent on a certain job and the peer-judgement.

Many other authors have pointed out the short falls and inadequacy of bibliometric indicators: (i) we have already noted that bibliometric indicators do not have rigorous theoretical backing; (ii) bibliometric indicators are required for their production, a bibliographic data-base which is to be fabricated according to the need. But these data bases are not readily available. The available data-bases have their own idiosyncrasies and may not be suitable for the purpose. Other data-bases are still to be manually prepared and are usually very small. There is always a question mark about the compatibility and sufficiency of the data, (iii) there is not yet any standard procedure for sampling bibliometric data.

These difficulties can be overcome if- (i) only those bibliometric indicators are used which are either straight forward or theoretically sound, (ii) machine readable bibliometric data-base are systematically produced as suggested by Sen in his proposal of ABSCINDEX [7] and by Isao Asai in his proposal of refraction data-base [8]; (iii) some standardization is evolved for comparing bibliometric data and for the optimum characteristic of bibliometric data-bases.

References

1. Sen, Subir K. and Chatterjee. Sunil Jumar. An introduction to researches in bibliometrics Part I. *Iaslic Bulletin* 35(3) 1990.
2. Vinkler P. An attempt of surveying and classifying bibliometric indicators for scientometric purposes. *Scientometrics* 13(5-6) 1988, 239-259.
3. Leydesdorff, L., and O. Amsterdamska. Dimensions of Citation Analysis. *Science, Technology and Human Values* 15, 1990, 305-35.
4. Diodato, V. *Dictionary of bibliometrics*, New York, The Haworth Press, 1994.
5. Steen, J C G Van. The role of science and technology indicators in science policy: do they really matter? Paper presented at the 4th International Conference on S&T Indicators, Autwerp, 5-7 October 1995
6. Sengupta, I N. Three new bibliometric parameters and physiology periodicals. *Annals Lib. Sci. Doc.* 35(3) 1988, 124-127
7. Sen, Subir K. ABSCOMDEX: Proposal of a new tool for information retrieval and science indicators in regional contexts. *Annals Lib Sci Doc* 35(3) 1988, 116-123.
8. Asai, Isao "Refraction" database for document information analysis (a report)