Datametrics? About the architecture of the metric disciplines

Victor Cavaller[1]

30 May 2008

Abstract

The community relationship maintained by the metric disciplines is expressed in concrete terms in three elements: the consideration of information as its object of study; the statistical and mathematical methodology that provides the basis of the scientific analysis techniques; and the specific subject or matter that metrics are developing.

We understand that the consideration of these elements is those that determine the architecture of the metric disciplines.

Metrics are defined as a combined whole of quantitative techniques associated with a specialised field of scientific knowledge, in order to obtain descriptive, evaluative or prospective results from the activities or phenomena analysis.

But some factors related to object, methodology and subject orientation of metric disciplines, cause that is unclear the position of informetrics in this architecture.

The aim of this paper is to examine the current debate of metrics architecture, to review some factors that condition informetrics status, and finally, to propose datametrics, as the macro-discipline where are integrated all metric disciplines.

1 Metric disciplines according to their object of study

The debate on the relationship between the metric disciplines is related to the acceptance afforded to the concept of information science.

We can define information science as the science aimed at the production, gathering, organisation, interpretation, storage, retrieval, transmission, dissemination, processing and use of information (Griffith, 1980). However, we agree with Capurro and Hjørland (2003) that "we have to place the role of an autonomous information science at a more abstract level, for which an epistemological reflection is needed that shows the difference between the concept of information for information science with respect to the use and definition of information in other sciences. A unified information science will integrate the considerations on the analogous, equivocal and univocal relationships between different concepts of information and their respective theories and fields of application."

In the field of information science, the metric and statistical analysis of information is the quantitative analysis of informational phenomena and is defined as informetrics. This was determined by the Informetrics Committee of the FID (1980), with the support of such leading authors like Gorkova, Brookes, Egghe, Rousseau, Hood and Wilson, etc. According to this criterion, bibliometrics and scientometrics are established as sub-fields of informetrics, which

[1]Open University of Catalonia, Dept of Information and Communication Sciences. Barcelona (Catalonia-Spain), vcavaller@uoc.edu

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

is defined as the study of the quantitative aspects of information (Brookes, 1990) and as a set of techniques for the application of mathematical methods and models to the study of the phenomena of informative scientific activity, with the aim of determining the structure and properties (not the content) of information, through the definition of the regularities of the communication processes of this information (Setién and Gorbea, 1990).

It is a criterion adopted by the majority of the scientific community. However, it should be understood, on the one hand, that this definition is restricted to the field of informational phenomena, and, on the other, that with regard to the relationship between metrics, the criteria that determine the definition of their competence should be defined.

From a point of view that goes beyond informational phenomena, the classification criteria of the metrics could be, for example: formal, such as webmetrics, to be competent on information in the web environment; relating to the field of study, such as scientometrics, to be competent on those factors that condition scientific production; or content, if we heed the specific nature of the variables that describe their object. For example: econometrics, which is defined to consider variables of an economic nature.

2 Methodology and techniques of information analysis procedures

With regard to the methodology and the techniques of information analysis procedures, in every science, scientific nature is expressed in the capacity of prediction, which refers to the quantitative aspect and the possibility of importing mathematical models from the sciences. To this effect, we should distinguish various fields that co-exist in information science:

- The discipline that defines "the set of techniques or activities related with the computerised treatment of data" that concerns Computer Sciences.
- The dimension relating to the procedural and interpretation method of the data in numerical value and of the information that concerns information statistics and mathematics.
- The discipline or sub-field of information science that deals with the quantitative aspects of information that concerns the different metrics: bibliometrics, scientometrics, etc.

Statistics uses the procedures and techniques used to collect, present, analyse and interpret data (Novales, 1997) to make inferences and on which to base decisions in a situation of uncertainty or when faced with incomplete information in the field of economics and other social and physical sciences (Salvatore, 1982). In its application to the field of information, statistics helps highlight the relationship between the variations of informational unities. Statistics is the set of techniques of mathematical interpretation applied to the analysis of numerical values when those applied to the values for which an exhaustive study is impossible due to their large number and complexity.

Mathematics, which is defined as the science that studies the properties of numbers, figures, sets, operations, functions, etc. when it refers to the analysis of ordered series determined by a calculation of informational objects, concludes with the function that claims to authenticate the mathematical regularity between variables that explains their distribution and enables a prediction to be made.

Metric disciplines have a methodological relationship with computer sciences and mathematics and statistic progress.

It is not surprising that Neter (1990) placed in the interaction of statistics with other sciences the impetus for the birth of new specialist informetrical technical branches. Information science is mathematised to enjoy the category of a science. Its concepts and laws are linked by formal relationships (Lafouge, Le Coadic and Michel, 2002). For example, the Bradford laws on the dissemination of scientific literature, Lotka's on the productivity of authors, constitute the determination of a rule, formula or relationship that is satisfied by the values of a vari-

able, the terms of a series of informational objects.

Together with the methodological factor that conditions the evolution of the metric disciplines, we should consider the phenomenon of ICT and the specialisation of scientific knowledge broadly studied from different viewpoints in the field of the social sciences and that, since the mid nineteenth century, have meant the appearance of new disciplines and branches of knowledge. The path of scientific progress has become a hybrid multi-specialisation (Dogan, 1997) and with the emergence of ICT we now have to add to it the new informational paradigm supported by a new technological paradigm based on the "growing and auto-multiplied data processing capacity in terms both of volume and of complexity and speed", on the capacity of recombination and on flexibility in the distribution of the new information technologies (Castells, 2003)

3 Architecture of metric disciplines

The expansion and deployment of new metric disciplines of information has been fostered from:

- The development of analytical techniques that statistics and mathematics offer to information science, and which give it validity as a science,
- The growing extension of the application of these quantitative analysis techniques to all the fields of human activity,
- The progress in the computation of the data that has universalised the use and application of these techniques,
- The phenomenon of specialisation of scientific knowledge,

The set of these factors accentuates the difficulties in establishing a detailed and exhaustive classification of all the potentially developable metric disciplines.

We can estimate that the factor that determines the prefix of the metric discipline comes from the subject orientation to which the information refers, from the nature of the information that it analyses or from the associated form of the information to be analysed.

However, going from the consideration of one definition criterion to another can lead to the scission of a metric discipline and generate another complementary one when experts of the latter expand or detail their competencies. So, for example, although the aim of bibliometrics is the quantitative study of publications and is, in this sense, distinguished by their object and form (printed communication), the aim of scientometrics is the quantitative study of the subset of these publications, which represent a scientific or technical activity (primarily articles and patents), but also all of the information that may provide knowledge of the scientific activity, which leads us to a criterion of field.

In other cases, it may cause the introduction of a sub-classification. We can also find examples outside the field of information science. In econometrics, we distinguish theoretical econometrics, which is defined by its object, i.e. by treating data of the relationships of an economic nature irrespective of the field in which these occur, of the applied econometrics that refer to the information of specific fields of the economy such as the ones mentioned above: theory of demand, production, investment, etc.

As we stated above, with regard to the architecture of the metric disciplines in the field of information science, there has been a partially resolved debate. There have been attempts to make informetrics a complementary discipline of an order parallel to bibliometrics and scientometrics. In this sense, in a comparative analysis McGrath identified the object of study, the variables, the methods and the objectives of bibliometrics, scientometrics and informetrics.

In the scientific literature, as concluded by Peres (2002) and Gorbea (2005), in line with the analysis of Gorkova (1988), Egghe and Rousseau (1990), Brookes (1990), Tague-Sutcliffe (1992), Russell (1994) and Hood and Wilson (2001), informetrics is considered as the set of the bibliometric, scientometric and webmetric metric disciplines aimed at the quantitative study of informational phenomena.

4 Proposal

We understand that the solution of restricting the field of action of informetrics to informational phenomena (publications, web, museums, media, etc.) may solve the problem in the field of information science, heeding the treatment and use of information as a resource. However, it leaves undone an essential objective of information science: the generation and structuring of information that goes beyond its form or the physical resource that presents it. If we consider information, in your current formal sense, as the object of quantitative analysis by informetrics into informational environments, what do we do about the other metric disciplines that are also aimed at the quantitative analysis of information? In other words, what is the relationship of informetrics with the other more popular metric disciplines such as econometrics, sociometrics, psychometrics, chemometrics and biometrics?

The metric disciplines have in common the application of statistical and mathematical techniques to their fields, but each one is defined complementing this application with expert knowledge on the subject. This knowledge appears in the creation of the specific indicators of application for each field.

If we preserve the definition of informetrics for the field of informational phenomena, what name is given to the macro-discipline that groups together the set of all the metric disciplines and which is aimed at the quantitative analysis of information irrespective of the field where it is generated?

In our understanding, the meaning – that information provides to data – determines the field of its application, for which reason, if we want to obviate the field of the different metric disciplines, we have to refer to a discipline that is aimed at the metric analysis of data.

In this sense, there are two solutions:
a) We can denominate informetrics to quantitative analysis of information beyond informational phenomena. Then, we must expand the meaning of informetrics in a non-formal acceptance towards a discipline that adds to the set of statistical and mathematical techniques, not an expert knowledge in a particular subject, but a common methodology for the systematic analysis of information based on an arrangement and classification of indicators. If we choose this solution, we have to extend its scope beyond informational phenomena and obviate formal demarcation criteria.

b) If, on the contrary, we preserve the term of informetrics for informational environments, we have to define a metric macro-discipline par excellence. We propose the term *datametrics* (Cavaller, 2006) to define this metric macro-discipline, which is defined as the set of analysis techniques aimed at information, irrespective of the form and field in which it is presented.

5 Discussion

Datametrics groups together in one single methodological body the set of metric sub-disciplines and is distinguished from the other metric disciplines in that it deals with the quantitative analysis of information without referring to any scientific field of specific application.

Datametrics is the quantitative analysis of information without formal considerations; it should be defined precisely as a process of formalisation of information. If it is metrics of data, to distinguish it from statistics and mathematics, it must provide a common methodology for the creation, arrangement and classification of indicators, and tecniques selection, aimed at the systematic analysis of information.

This is the challenge facing us. From the consideration of the systematic dimension of the analysis of information, we can now formulate the question that becomes key for us: what are the *modes* of analysis that highlight the groupings of indicators in the different disciplines of metric analysis associated with the different types of monitoring in organisations? This and other questions must be resolved in subsequent studies (Cavaller, 2008).

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

References

Almind, T.C.; Ingwersen, P. (1997). "Informetric analysis on the World Wide Web: methodological approaches to 'webometrics'". *Journal of documentation*, 53 (4).

Araújo, J.A.; Arencibia, J. (2002). "Informetría, bibliometría y cienciometría: aspectos teórico-prácticos". *Acimed*, 10 (4). [Consulted: 17/03/2008]. Available at: <http://www.infomed.sld.cu/revistas/aci/vol10_4_02/aci040402.htm>.

Bradford, S.C. (1934). "Sources of information on specific subjects". *Engineering: an ilustrated wekly journal*, 137, 85-86.

Brookes, B.C. (1990). "Biblio-, scionto-, inform-metrics??? What are talking about?". En: Egghe i R. Rousseau (eds.). *Informetrics 89/90. Selection of papers submitted for second International Conference on bibliometrics, scientometrics and informetrics*. Amsterdam: Elsevier Science Publishers.

Callon, M.; Courtial, J.P.; Penan, H. (1993). *La Scientométrie*. Que sais-je?, nº 2727. Paris: Presses Universitaires de France.

Capurro, R. (2003). "Epistemología y ciencia de la información". En: *V Encontro Nacional de Pesquisa em Ciência da Informação*, Belo Horizonte (Brasil), 10-12 nov. [Consulted: 01/03/2008]. Available at: http://www.capurro.de/enancib.htm

Capurro R.; Hjørland, B. (2003). "The Concept of Information". *Annual review of information science and technology*, 37, 343-411.

Cavaller, V. (2006). ""Matrix indicator system for strategic information analysis: application in a practical case". Dissertation: University of Barcelona. Available at: http://www.tesisenxarxa.net/TDX-0326107-111907/index.html

Cavaller, V. (2008). "Matrix Indicator System". In: *Proceedings of Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting*. Humboldt-Universität zu Berlin. Institute of Library and Information Science. Berlin (Germany), 29 July–1 August.

Cole, F.J.; Eales, N.B. (1917). "The history of comparative anatomy". *Science progress*, 11, 578-596.

Cronin, B.; McKim, G. (1996a). "Markets, competition and intelligence on the World WideWeb". *Competitive intelligence review*, 7 (1).

Cronin, B.; Mckim, G. (1996b). "Science and scholarship on the World Wide Web: a North American perspective". *Journal of documentation*, 52, 163-171.

Egghe, L.; Rousseau, R. (2002). "A general framework for relative impact indicators". *Canadian journal of information and library science/Revue canadienne des sciences de l'information et de bibliothéconomie*, 27 (1), 29-48.

Elkana, Y.; Lederberg, J.; Merton, R.K.; Tackray, A.; Zuckerman, H. (1978). *Toward a metric of science: the advent of science indicators*. New York: John Wiley & Sons.

Gorbea, S. (2005). *Modelo teórico para el estudio métrico de la información documental*. Gijón: Ediciones Trea.

Gorkova, V.I. (1988). "Informetrics". *Informatics*, 10. Moscú: VINITI.

Griffith, B.C. (1980). *Key papers in information science*. New York: Knowledge Industry Publ.

Gross, P.L.K.; Gross, E.M. (1927). "College libraries and chemical education". *Science*, 66, 385-389.

Guzmán, M. V. (1999). *Patentometría: herramienta para el análisis de oportunidades tecnológicas*. Tesis de doctorado. Gerencia de información tecnológica. Universidad de la Habana: Facultad de Economía.

Hood, W.W.; Wilson, C.S. (2001). "The literature of bibliometrics, scientometrics and informetrics". *Scientometrics*, 521 (2).

Kalyane, V. L.; Sagar, A.; Kumar, A.; Kumar, V.; Mohan, L.; Prakasan, E. R. (2003) "Librametric mapping of the 'libraries, archives & information technology' R &

D during 1970-1990". *E-LIS: E-Prints in library and information science*. [Consulted: 01/03/2008]. Available at: <http://eprints.rclis.org/archive/00001039/01/Librametrics_PDF.pdf>.

Lafouge, T.; Le Coadic,Y.F.; Michel, Ch. (2002). *Eléments de statistique et de mathématique de: infométrie, bibliométrie, médiamétrie, scientométrie muséométrie, webométrie: cours avec exemples et exercices corrigés*. Villeurbanne: Presses de l'Enssib, École supérieure des sciences de l'information et bibliothèques.

Leydesdorff, L. (1989). "The Relations Between Qualitative Theory and Scientometric Methods in Science and Technology Studies". *Scientometrics*, 15 (5-6).

Morales, M.; Cruz, A. (1995). "La bibliotecología, la cienciología y la ciencia de la información y sus disciplinas instrumentales: su alcance conceptual". En: *Ciencia de la Información*, 26 (2), 70-88.

Novales, A. (1997). *Estadística y econometría*. Madrid: McGraw-Hill.

Peres, N.A. (2002). "Da bibliometria à webometria: uma exploraçao conceitual dos mecanismos utilizados para medir o registro da informaçao e a difusao do conhecimento". *Ci.Inf., Brasília,* 31 (2), 152-162.

Polanco, X. (1997). *Análisis de la información y tecnologías de la inteligencia: Contribución de la inteligencia artificial al estudio de la ciencia y la tecnologia.*. RICYT: Biblioteca. [Consulted: 01/03/2008]. Available at: <http://www.ricyt.edu.ar/interior/biblioteca/outlook.pdf.>.

Pritchard, A. (1969). "Statistical bibliography or bibliometrics?". *Journal of documentation*, 25 (4), 348-349.

Rip A.; Courtial J-P. (1984). "Co-words maps of biotechnology: an example of cognitive scientometrics". *Scientometrics*, 6 (6).

Russell, J.M. (1994). "Back to the future for informetrics?". *Scientometrics*, 30 (1), 407-410.

Salvatore, D. (1982). *Econometría. Problemas resueltos*. Madrid: McGraw-Hill.

Smith, A.G. (1999). "The impact of web sites: a comparison between Australasia and Latin America". En: *Proceedings of INFO'99. Congreso Internacional de Informacion*, La Habana, 4-8 Octubre. [Consulted: 01/03/2008]. Available at: <http://www.vuw.ac.nz/staff/alastair_smith/publns/austlat/>.

Spinak, E. (1996) Diccionario enciclopédico de bibliometría, cienciometría e informetría. Caracas: Unesco.

Tague-Sutcliffe, J. (1992). "An introduction to informetrics". *Information processing & Management,* 28 (1), 1-3.

Van Raan, A.F.J. (1997). "Scientometrics: state of the art". *Scientometrics,* 38 (1), 205-218. [Consulted: 01/03/2008]. Available at: <http://www.cwts.nl/TvR/documents/AvR-StateArt-Sciento.pdf>.