Jointly published by Kluwer Academic Publishers, Dordrecht and Akadémiai Kiadó, Budapest

Scientometrics, Vol. 50, No. 1 (2001) 59–63

Bibliometrics and internet: Some observations and expectations*

ANTHONY F. J. VAN RAAN

Centre for Science and Technology Studies (CWTS), University of Leiden, Leiden (The Netherlands)

Electronic publishing developments and new information technology in general will affect the main functions of scientific communication. Most changes however will be primarily technological but not conceptual. Publication via journals of high reputation is in most fields of science crucial to receive professional recognition. That will remain so in the "electronic era". A much more revolutionary change in science will be the increasing availability and sharing of research data.

Introduction

In this paper we elaborate on our earlier work on the role of the quality assurance system in science in the era of electronic publishing (*Van Raan,* 1997). The active scientist is both an author and a reader of publications. *Roosendaal* (1995) clearly formulates this dichotomy in the scientific communication process. As an author, the scientist is knowledge-dissemination driven. He or she wants to publish as much as possible. But as a reader the scientist is knowledge-concentration and -integration driven, wanting to read as less as possible. This dichotomy is important in the discussion about scientific communication in general and about the electronic developments in particular.

The electronic developments undoubtedly enable further growth in documented knowledge. This is generally regarded as positively related to accessibility of knowledge and openness of research results. However, the rate of growth of scientific publishing can also be seen as a threat for the progress of science. Many scientists already observe an overload of published material of mediocre value, and fear that the internet will makes things worse. Growth also enables ever-increasing specialization leading to further differentiation or, as some maintain, further fragmentation of science. However, the electronic explosion of information and documents can also provide new means of accessibility, integration, and discovery of so far hidden "knowledge patterns" (*Noyons* and *van Raan,* 1998).

* Based in part on an original article by the author (*van Raan*, 1997) used with the kind permission of Reed Business Information.

0138–9130/2001/US $ 15.00

Main functions of scientific communication

Physicists in the field of string theory (elementary particle physics) maintain that they even do not need the famous journal *The Physical Review* for their communications (see *Langer,* 2000) and that the Los Alamos e-print archive (http://xxx.lanl.gov) sufficiently covers all their interests. Most other scientists, also in physics, will most probably not follow this way of life as it inevitably will lead to a kind of scientific "autism". Furthermore, *Poultney* (1996) observes strong reservation, often because of the danger of plagiarism. So we can be sure that electronic publishing developments will evolve at a different pace and in different forms within science, depending on fields and research specialties.

Given the important role of citation behaviour of scientists in the recognition system, what does the near future hold for citation practices and citation-based bibliometric methods in relation to the electronic developments? Some people suggest that electronic document facilities such as hypertext-linking will replace largely the "classical" way of citing. This is very unlikely. Hypertext-linkage as new type of citation is too anonymous, too superficial. It does not provide a clear recording of the use of other's work. Furthermore, there are conceptual differences between hyperlinks and classical citations, particularly in terms of a meaningful quantitative assessment of the "impact" (*Egghe,* 2000). Therefore hypertext-linking fails as a serious element in the rewarding process, which is crucial in the reputation system of science. Hypertext-linking will certainly explode as a tool, like fax and e-mail, but essentially not much more than that.

We expect that the electronic developments will not replace classical citation behaviour but, quite the contrary, will enhance it. Electronic versions of journals will more and more provide facilities to directly link to articles mentioned in the reference list or to delivery services if the cited articles cannot be made directly available in an electronic way. We foresee that by these developments classical citation behaviour will achieve a higher level of activity and thus will improve bibliometric analysis considerably. Moreover, we expect that an increased level of hyperlinking between scientific publications via references will lead to an increase in the average age of citations.

The scientific journal will not disappear. The electronic version will simply be added to the paper version and will become rapidly more important, mainly because of many additional facilities like the one mentioned of above. Why will the journal survive? We must realize that the concept of a scientific journal embodies all main functions that are essential in scientific communication. According to *Roosendaal* (1996) these main functions are:

(1) certification, which directly relates to quality assurance according to scientific standards; (2) registration, which directly relates to ownership protection and reward; (3) awareness, directly related to disclosure and search of knowledge; and (4) the archival function which relates to storage, accessibility, and to new techniques called "knowledge discovery" and "data mining".

The crucial role of the scientific archive

The above, however, does not mean that these four main communication functions as can only be fulfilled by journals. What are the possible electronic non-journal alternatives? Take for instance the first main function: certification and its direct relation to quality assurance. Although peer review plays a central role in the scientific endeavour, there is also the wish to get rid of the established referee (peer review) system for publication selection in journals and to have a more free dissemination of knowledge. Quality control, optimists believe, could be achieved in a more "democratic" way: comments of colleagues will improve the electronically posted papers, manifestly bad or even misleading work will be pilloried and ignored.

Poultney (1996) states that most probably a "two-tier system" will evolve. The first tier is a "free space" for preprints and other "preliminary" communications. It is a representation of the scientific enterprise in "real time". The second tier is the world of the more formal publications. The most important role of the first tier will be that original work cannot be torpedoed by a publication veto (in the framework of formal journal refereeing) of competing groups. This first tier is the place where informal and formal communication will become closely connected. Readers will take on the role as "informal referees". Thus, in the first tier the above "democratic" review of posted papers will arise. An important point is the earlier mentioned fear for "plagiarism" and "theft of ideas". This means that already the first tier facilities to record "intellectual property" must exist. If this first tier is sufficiently "stable", documents can be cited by an archive number.

Formal submission of papers takes place in the second tier. In many cases this will be a "promotion" from the first tier, by using a limited number of standardized formats. Decisions of the referees in the second tier may or may not follow the "first tier" opinions. It will be a more selective quality control. This second tier is particularly important for all other main functions of scientific communication such as registration related to ownership protection, awareness based on a standardized disclosure of research results, and the archival function. Most probably, this second tier will consist largely of the already existing journals, of which most soon will have a printed as well as an electronic form.

A further important function of the second tier with all its journals is the archival role. But as probably not all documents in the first tier will "promote" to the second, the archival work will be a joint concern of both the organisations maintaining the first tier and those involved in the second tier. Long-term archiving of electronic publications is still a major physical problem due the relatively short lifetime of electronic information carriers as compared to paper. This problem must be solved: the archival world is very important. In many fields of science publications of thirty or more years old can still be relevant for current work at the scientific frontiers. Similarly, work published today might be relevant for science thirty years from now. Our recent work on growth of science shows that a hard core of scientific literature older than about ten years "ages" considerably slower than the more recent literature (*van Raan,* 2000a). So we simply have the duty for our future colleagues to archive our published work thoughtfully (*Langer,* 2000). Moreover, the archive offers a new and strong potential: in our recorded human knowledge much further knowledge is "hidden". Data-mining and "knowledge discovery" techniques will enable us to explore unexpected patterns and linkages, particularly across scientific disciplines. These developments will facilitate the scientists in his or her role as a "reader" particularly in "knowledge integration" activities.

Another crucial step will be the linkage of this publication archive to special research databases in order to enable the exchange and collective processing of very large volumes of data. Scientists, and again predominantly physicists, are now experimenting with special data sharing networks or "data grids". These "new style webs" (*Voss,* 2000) are a very important new and rapid development, particularly in fields such as high energy physics, genome research, and astronomy where huge instruments produce so many data that sharing the research results with a much larger group of scientists than only those directly involved in the research projects concerned becomes a necessity.

Concluding remarks

For a reader of a scientific document it will remain important to know whether a given citation concerns work of sufficient quality. A journal provides a certain quality assurance. So the most important role of editors and referees remains guarding the reputation of a journal, and with that the reputation of all the authors who succeed in publishing their articles in that journal. Scientific communication and reputation are strongly linked via journal-status (*van Raan* 2000b). Almost nothing in the scientific enterprise can compete with the importance of a publication in top-journals. It, in fact, constitutes the main source of recognition. Thus, all new electronic publication systems

will have to compete with existing journals – be it in electronic form – for their ranking in scientific status. One way or another, citation behaviour of scientists will remain essential in the determination of this status.

Bibliometrics and the Internet: Plus ça change, plus c'est la même chose. "Real time" web-based reporting and commenting about research results will be not a "replacing" but an additional facility in the whole of scientific communication. A much more revolutionary change in science will be the increasing availability and sharing of research data, i.e., the emergence of a real time web-based collective use of research material.

References

L. EGGHE (2000). New informetric aspects of the Internet: some reflections – many problems. *Journal of Information Science,* 26 (5) 329–335.

J. LANGER (2000). Do string theorists need the Physical Review? *Physics Today,* 53 (8) 35–38.

E. C. M. NOYONS, A. F. J. VAN RAAN (1998). Monitoring scientific developments from a dynamic perspective. Self-organized structuring to map neural network research. *Journal of the American Society for Information Science (JASIS),* 49 (1) 68–81.

R. W. POULTNEY (1996). Front-Ends are the way to go. *Europhysics News,* 27 :24–25.

A. F. J. VAN RAAN (1997). The future of the quality assurance system. Its impact on the social and professional recognition of scientists in the era of electronic publishing. *Journal of Information Science,* 23 (6) 445–450.

A. F. J. VAN RAAN (2000a). On growth, ageing and fractal differentiation of science. *Scientometrics,* 47 (2) 347–362.

A. F. J. VAN RAAN (2000b). Competition amongst scientists for publication status. Toward a model of scientific publication and citation distributions. *To be published.*

H. E. ROOSENDAAL (1995). Roles of bibliometrics in scientific communication. *Research Evaluation,* 5 :237-241.

H. E. ROOSENDAAL (1996). Scientific Communication and the Science Process. *Proceedings of the International Summer School on the Digital Library*. Tilburg: TICER.

D. VOSS (2000). Data networks: scientists weave new-style webs to tame the information glut, *Science,* 289 (5488) 2250–2251.

Received November 10, 2000.

Address for correspondence:
ANTHONY F. J. VAN RAAN
Centre for Science and Technology Studies (CWTS),
University of Leiden, Wassenaarseweg 52
P.O. Box 9555, 2300 RB Leiden (The Netherlands)
E-mail: vanraan@cwts.leidenuniv.nl