Inserm

Institut national
de la santé et de la recherche médicale

Draft project on « Electronic Publication » of the *Comptes Rendus de l'Académie des Sciences*

Philippe Kourilsky, Suzy Mouchet and Colette Brézin

Paris – February 1992

I. BACKGROUND

Many scientists feel a growing malaise over problems linked to communication and publication, and their corollaries, namely the evaluation of personnel, teams and projects. The considerable increase in the volume of publications in the Life Sciences and health matters is characteristic of this branch of scientific investigation: between 10 and 100 times more pages are published each year in biology than in chemistry or physics, for example. This reflects not only the high intrinsic « productivity » in this field, but also the increase in the number of researchers in biological disciplines: owing to the deep structural unity of living organisms, studies of *Drosophila* or begonias can be relevant to our understanding of human biology.

This increase in the number of publications means that individual scientists are finding it harder to keep abreast of scientific information, even in a given scientific domain, and documentary watch activities are increasingly becoming a team effort. No one can possibly read in detail a dozen journals, and this tends to reinforce the hegemony of major international journals such as *Nature, Cell,* etc. The abundance of documentation is such that many scientists merely skim through journals and, as a result, gain only an overview, by reading the summaries and titles or, often, the titles alone. Many researchers gather piles of photocopies of articles they consider important but that they have only skimmed through. This accumulation provides a false sense of security regarding knowledge on a particular subject. One consequence is the increasing importance of oral communication, whereby scientists glean « predigested » information and save considerable time on reading. In addition, oral communication well adapted to the team situation. It thus presents many advantages, but it also suffers from a certain subjectiveness: rumors, researchers' renown, etc. are thus conveyed orally, but also in highly random fashion (chance meetings, conferences, seminaries, etc.).

One might think that increasing specialization would suffice to deal with a problem which, in essence, is individual*, i.e.* that depends on the capacity of individuals to solve it. This is not so: the decrease in the validation of information has far more perverse consequences, because it affects the quality of evaluations. Errors and a lack of precision are increasingly frequent during anonymous peer review of articles submitted for publication. Referees no longer have the time to analyze manuscripts thoroughly, and they do not always have the necessary expertise. No one now checks the validity of cited articles. Because of a lack of space, authors and publishers implicitly agree to reduce the number of references and simply cite review articles instead. The latter, unfortunately, are all too often opinionated or, on the contrary, simple non critical compilations of data from several hundred articles. Too rare is the honest, critical and well-balanced review. Manuscript review by anonymous peers only leads to the rejection of a marginal number of articles: an article rejected by one journal almost always ends up being published in another. Peer review, in contrast, improves the quality of publications, provided that referees take this important scientific role seriously.

And what is really at stake in peer review of original manuscripts is the evaluation of individual researchers, teams and projects. When evaluating research teams and projects, specialized commissions and scientific advisory boards take into account, at least to a certain degree, the quality of papers reflected by the « impact factor » of the journals in which they were published.

The notoriety of the journal impacts on the perceived quality of both the article and its authors. The major reviews generally publish only « good » articles, but not all « good « articles are published in the most prestigious journals, as the choice of manuscripts for publication can depend partly on non scientific criteria, such as trends, authors' own notoriety, networks of colleagues, editorial policies, etc. Thus, the assessment of researchers, teams and projects is partly based on systems that are known to be imperfect (and sometimes partial) but which, nonetheless, still have considerable influence.

Before leaving this aspect of the problem, one must mention the cultural problems inherent in a system of publication that is largely dominated by Anglo-Saxons, and especially the Americans. The use of English to convey original research findings is, for the moment, a *fait accompli*, and this engenders constraints on thought processes, writing and communication in general by those whose first language is not English. Beyond purely linguistic problems, cultural traits (for example, « speculation » is frowned on in the United States) and sociological characteristics (« publish or perish » in the current American system and others) impacts strongly on our own values. This has a number of positive aspects, helping for example to overcome certain national weaknesses, but it also inflicts upon us certain weaknesses inherent in other cultures. It is perfectly justified to resist these phenomena when necessary.

Regarding the evolution of the current system, and leaving to one side the new tools that will be discussed below, we see no reason for optimism. On the contrary. It seems that, over the last decade, the gradual loss of efficiency in the transmission of information, together with the degradation in the validation of information, has gone hand in hand with a certain loosening of scientific values and a growing number of « slight dishonesties ». What matters therefore, first and foremost, is that the scientific community should again master information and its communication. This would lead to improvements in research and its evaluation.

II. DEFINITION OF "SCIENTIFIC FACT"

The new methods of communication based on the use of computers and other electronic tools offer a number of potential solutions to those problems. However, we would first like to introduce an extremely important notion: these new techniques should not simply serve to handle a growing number of publications regardless of their contents. And rather than thinking it terms of the « utility » of scientific papers, we must come back to the basic notion of scientific fact. Put simply, every publication should contain new scientific fact. Is the current multiplication of publications due only to an increase in scientific knowledge, or is it also due, at least in part, to an impoverishment of the notion of « scientific fact ». Even if this is not the case, there might be a case of reinforcing the definition of « scientific facts warranting publication ». This recalls a fairly widespread elitist opinion that S. Tonegawa, for example, formulated as follows: each individual, in his or her research career, makes only a small number of important discoveries. Therefore, each researcher should be given 50 « tickets » authorizing publication of no more than 50 articles during his or her career. Yet the current trend is precisely the reverse: there are a growing number of small but highly specialized journals with limited but highly targeted readerships that rarely report major discoveries. New techniques, which have cut the cost of printed publications, have created a multitude of small markets for journals with print runs of only 700-1000, most copies being bought by libraries. Thus, the current reaction when faced with the difficulty of publishing in a « major » journal is not to revise the definition of scientific fact but to multiply the number of vectors.

It is therefore crucial to decide whether the new technologies will simply serve to manage the current proliferation of scientific papers, or whether their use will be accompanied by a redefinition or requalification of scientific fact.

III POSSIBILITIES OFFERED BY NEW TECHNOLOGIES

The two principal technological developments concerned here are computerized databases and communications systems (especially networks).

A) Databases produced from the printed page

The growth of computerized scientific databases is a relatively recent phenomenon, which will probably have important implications for the problems outlined above.

At present, the titles of all articles published in almost all scientific reviews worldwide (indexed by *Current Contents*, for example) are available in the form of diskettes sent out on a weekly basis, one to two weeks after publication. More recently, summaries of articles have become available, with a delay of a few weeks after publication. A database such as Medline offers monthly CD-ROMs containing both titles and summaries, but only appears between 3 and 6 months after publication. Other databases, including European and French databases, are available in various forms. Many are accessible on line, but the cost is often prohibitive for individuals (rates are often adapted to use by a large number of readers consulting the databases in libraries).

Algorithm aimed at helping with efficient and « intelligent » database searches have progressed spectacularly. These improvements, together with growing microcomputer power, are making literature searches increasingly efficient.

In principle, the (probable) general adoption of computerized search methods will eventually reduce the impact of the major international journals, as an article published in a lesser journal will no longer be missed during literature searches. This hypothetical improvement will depend very closely on the performance of search software and/or the substructure of the databases themselves (we seem to be evolving towards specialized databases that are simpler to consult than general databases). Indeed, key-word searches often yield hundreds of references and summaries, and the person is thus required to read the equivalent of a book, almost bringing us back to the pre-existing problem. In addition, the information is provided (at present in the form of summaries) in unclassified manner, meaning that readers may still, without any objective reason, consider that an article published in *Nature* is better than one published in a more specialized review. At least the latter will no longer be simply overlooked.
Despite these reservations, the general use of computerized databases is clearly a major step forward.

B) Improving network communication

The success of the fax exemplifies how rapidly networked communication can change working habits. Yet the fax is only a primitive form of electronic communication. In some fields (especially physics and computing), electronic mail predominates over all other forms of communication. E-mail means that communication is no longer a simple dialogue between two individuals, but communication between *n* individuals forming a network.

Based on the experience acquired over the last decade by an international group of theoretical physicists, and on the current debate over purely electronic publication (which will be discussed below), we can outline the following scenario: groups of scientists whose activities are centered on a relatively precise theme can communicate electronically, within a network, far more rapidly and effectively than by any other means. This technological advance introduces a radical change in the rules of communication. New information is injected into the network with no preliminary controls and well before its publication in printed form. Publication on the electronic network « patents » the information on that date. The network serves, by the reactions it elicits, as a group of referees, meaning that the assessment of the new information placed on the network is in fact a collective process. Depending on how the author's work evolves, and on the reactions of the network, the initial report can be changed, and this occurs « in public ». When the work is judged to be ripe, it can be sent to a journal for publication in print, sometimes more than a year after it first appeared on screen. This resembles the system whereby, before submitting it for publication, an author sends his or her work to a club of colleagues for their reactions. This system, which was in vogue among biologists in the 1970s, has more or less disappeared but is starting to re-emerge.

Here are some characteristics of electronic messaging

1) The information is available instantaneously worldwide in centers equipped with e-mail facilities.
2) The information is available to laboratory managers, researchers and students alike, with no censoring or distortion.
3) Simple microcomputers can serve as terminals connected to a local network. The communication server that pilots the system for a group of laboratories at a given site is a shared workstation; the cost of these computers is constantly declining (currently about ten thousand dollars).
4) The information is cheap, and places on an (almost) equal footing both large and small institutions, in the developing and industrialized countries alike.
5) Most universities worldwide already have such communication networks (mainly instigated by physicists).
6) Computer storage capacities are considerable, even if the amount of information is enormous (cf. computations by the physicist P. Ginsparg)
7) Nevertheless, oral communication will remain crucial, especially to build bridges between hyperspecialties. Scientific journals devoted to review articles will almost certainly play an increasing rôle, and the review articles they publish should be taken into account during auditing procedures.

Several remarks can be made.

1) First, the system described above is a convivial working tool rather than an electronic « journal »
2) Second, if such networks are created in the life sciences and medicine they will have to be of relatively limited size, both because of technical limitations proper to the communication network (which may evolve with time) and because of the fact that, to be effective, such a network can only be provided by a relatively limited amount of information. Even if these conditions are respected it will be necessary to develop efficient search software.
3) Finally, as these networks must be of limited size to function correctly, a set of rules must be developed to avoid deviations and misconducts. For example, a network should be centered on an precise theme but be open to all. The opposite (a closed network) could lead to a club of a « happy few » with access to preferential information, seriously handicapping those who do not belong to the club (especially younger researchers). Finally, if a number of such networks are created, given individuals may participate in several. Furthermore, the networks should be interfaced and possibly be structured hierarchically into a « network of networks », etc.

IV THE POSSIBLE ROLE OF THE FRENCH ACADEMIE DES SCIENCES

As mentioned above (cf section II), one of the key points for the future is the possible redefinition, or requalification, of « scientific fact ». This question clearly exceeds the framework of INSERM, and can only be dealt with by national, European or international bodies. We propose that the Académie des Sciences approaches this question experimentally. This leads us to propose a collaborative pilot project between Inserm and the Académie des Sciences (which could, of course, be extended to include other organizations).

On the one hand the status of the Académie des Sciences is adapted to the scale of the problem and, on the other hand, the length of Notes submitted to *Comptes Rendus* is suited to the communication of new scientific facts, in keeping with their original vocation. The experiment, the main principles of which are described below (pending more thorough discussion) could help to reinvigorate *Comptes Rendus de l'Académie des Sciences* without involving radical changes. Finally, the academy could fulfill its educational rôle, in a spirit of innovation, and have a profound impact on the scientific community. Of course, it might be desirable for this project to be opened to other European partners.

V. A PROJECT FOR ELECTRONIC MAIL AND PUBLICATION

In brief, under the aegis of the Académie des Sciences, an electronic mail network could be set up with the following principles:

1) The network should be available to all, and be as simple and inexpensive as possible.
2) The network should be furnished by scientific notes whose length corresponds approximately to that of Notes currently submitted to CRAS (i.e. about 3 or 4 typewritten pages with a few figures or tables).
3) In principle, these Notes should contain new « scientific fact ». The note format will often be too short to report all the experimental data required to establish the validity of the new scientific fact. The author will then be obliged, by virtue of a clearly stated deontology, to supply complementary scientific data to any interested person on request.
4) The Notes will be transmitted to the academy and can belong to three consecutive categories designated A, B and C. Any note submitted to the academy starts in category C. It must be validated (length and general contents) before being upgraded to category B.
5) Admission to categorie B signifies that the note can be entered into the database without undergoing peer review. In other words, notes enter the database under the authors' responsibility. They then become accessible to all those interested.
6) The authors can request classification of the note in category A, which implies anonymous peer review. Readers of the note (accessible via the database) can also request classification of the note in category A if they judge the results to be important.
7) Notes in category A compose the subset of the database that is destined for publication in CRAS. The database itself contains an image of the national scientific patrimony.
8) Readers of notes contained in the database can, electronically, qualify them as belonging to category A or B. They can therefore contest a classification in category A or B, even several years after the note has been deposited in the database. Thus, important results that were initially underestimated can be recognized for their true value.
9) Notes should be submitted in both French and English. This should not be a major obstacle, as the notes are relatively short. If this is the case, the printed form of CRAS would have to exist in both French and English. The English database would clearly facilitate international access.

This project takes into account the problems raised at the beginning of this report. In particular, the gradual requalification of « scientific fact » ; will be dealt with collectively by the scientific community, under the Academy's moral authority. The multidisciplinary nature of the Academy (this project is meant to include all disciplines) would no doubt facilitate specific development of the Life Sciences. The efforts so far undertaken at Inserm to modernize scientific communication would allow this institution to participate in a project of this type concerning series III (Life Sciences) of the *Comptes Rendus de l'Academie des Sciences*.